UNIVERSIDAD PANAMERICANA –UPANA-
DEPARTAMENTO DE ESTUDIOS DE POSGRADOS
DOCTORADO EN INVESTIGACIÓN SOCIAL
INVESTIGACIÓN ESPECIALIZADA
DOCUMENTO No. 18

SIERRA BRAVO, RESTITUTO.: *TÉCNICAS DE INVESTIGACIÓNSOCIAL: Teoría y Ejercicios.* Madrid: Editorial Paraninfo. 2001., pp. 173-703

Técnicas de Investigación Social
TEORÍA Y EJERCICIOS

He aquí una obra muy apreciada por su alto valor pedagógico, que comprende no sólo una exposición básica, completa y sistemática de las Técnicas de Investigación Científica y Social actualmente vigentes, sino también una extensa colección de casos prácticos seguidos de sus respuestas, fundados en muchos casos en investigaciones efectivamente realizadas.

Los ejemplos prácticos pueden servir tanto para ejercitar los conocimientos teóricos aprendidos, como para su mejor comprensión y asimilación, al conocer, de forma más concreta, alguna de las aplicaciones que pueden tener en la realidad. Por su claridad, es apta para todo curso, sea o no de nivel universitario, sobre las Técnicas de Investigación Científica y Social y por los ejercicios puede ser muy útil como complemento de cualquier otro manual o tratado.

Especialmente recomendable para los alumnos que cursan estudios en sociología, ciencias políticas, psicología, ciencias de la información, economía, estadística y trabajo social, y, en general, para todos los interesados, bien sean profesores, alumnos o investigadores, en la investigación científica y en la social.

PARANINFO
THOMSON LEARNING

www.paraninfo.es
www.thomsonlearning.com

ISBN 84-283-2429-8
9 788428 324298

TÉCNICAS de INVESTIGACIÓN SOCIAL
Teoría y ejercicios

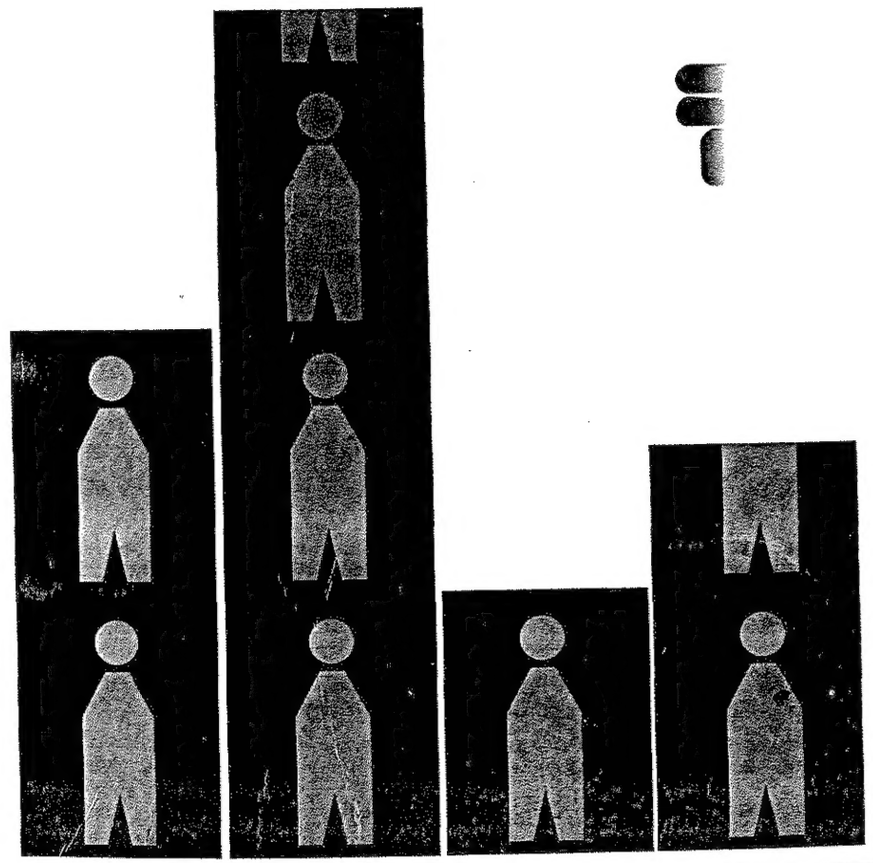
PARANINFO

PARANINFO
THOMSON LEARNING

Técnicas de Investigación Social
TEORÍA Y EJERCICIOS

DECIMOCUARTA EDICIÓN

R. Sierra Bravo



RESTITUTO SIERRA BRAVO

Técnicas de Investigación Social

TEORIA Y EJERCICIOS

DECIMOCUARTA EDICIÓN

PARANINFO
THOMSON LEARNING

Australia • Canada • Mexico • Singapur • España • Reino Unido • Estados Unidos

PARANINFO
THOMSON LEARNING

Técnicas de investigación social
© Restituto Sierra Bravo

Impresión:
Unigraf, S.L.
Móstoles - Madrid

COPYRIGHT © 2001 International Thomson Editores Spain
Paraninfo, S.A.
Magallanes, 25; 28015 Madrid
ESPAÑA
Teléfono: 91 4463350
Fax: 91 4456218
clientes@paraninfo.es
www.paraninfo.es

Impreso en España
Printed in Spain

ISBN: 84-283-2429-8
Depósito Legal: M-11.184-2001

(032/65/56)

Reservados los derechos para todos los paises de lengua española. De conformidad con lo dispuesto en el artículo 270 del Código Penal vigente, podran ser castigados con penas de multa y privación de libertad quienes reprodujeren o plagiaren, en todo o en parte, una obra literaria, artistica o cientifica fijada en cualquier tipo de soporte sin la preceptiva autorización. Ninguna parte de esta publicación, incluido el diseño de la cubierta, puede ser reproducida, almacenada o transmitida de ninguna forma, ni por ningún medio, sea éste electrónico, químico, mecanico, electro-óptico, grabación, fotocopia o cualquier otro, sin la previa autorización escrita por parte de la Editorial.

Otras delegaciones:

Mexico y Centroamerica
Tel. (525) 281-29-06
Fax (525) 281-28-66
clientes@mail.internet.com.mx
clientes@thomsonlearning.com.mx
Mexico, D.F.

Puerto Rico
Tel. (787) 758-75-60 y 81
Fax (787) 758-75-73
thomson@coqui.net
Hato Rey

Chile
Tel. (562) 531-26-47
Fax (562) 524-46-88
devoregr@netexpress.cl
Santiago

Costa Rica
EDISA
Tel./Fax (506) 235-89-55
edisacr@sol.racsa.co.cr
San José

Colombia
Tel. (571) 340-94-70
Fax (571) 340-94-75
clithomson@andinet.com
Bogota

Cono Sur
Tel. (5411) 4325-22-36
Fax (5411) 4328-18-29
thomson@thomsonlearning.com.ar
Buenos aires

Republica Dominicana
Caribbean Marketing Services
Tel. (809) 533-26-27
Fax (809) 533-18-82
cms@codetel.net.do

Bolivia
Librerias Asociadas, S.R.L.
Tel. Fax (591) 2244-53-09
libras@datacom-bo.net
La Paz

Venezuela
Ediciones Ramville
Tel. (582) 793-20-92 y 782 29-21
Fax (582) 793-65-66
telibros@artglobal.net
Caracas

El Salvador
The Bookshop, S.A. de C.V.
Tel. (503) 243-70-17
Fax (503) 243-12-90
amorales@sal.gbm.net
San Salvador

Guatemala
Textos, S.A.
Tel. (502) 368-01-48
Fax (502) 368-15-70
textos@infovia.com.gt
Guatemala

INDICE DE MATERIAS

Prefacio 11

Sección 1. – FUNDAMENTOS

1. Noción y fundamento de la asignatura 14

La ciencia (17). Los métodos de investigación. El método de investigación científico. Su noción y rasgos (20). El método científico en las ciencias sociales (23). Clases de métodos de las ciencias sociales (25). Método científico y técnicas científicas (26).

2. La investigación social 27

Peculiaridades de la investigación social y sus problemas (28). El problema de los valores en la investigación social (30). Tipos de investigación social (32). Ejercicios (37).

3. El proceso de la investigación social 41

Los aspectos del proceso de la investigación científica (41). El proceso metodológico (42). Operaciones de la investigación social (43). El proceso lógico (45). El proceso de verificación. Teoría (46). Modelos (48). Hipótesis y hechos. La verificación: su noción y aspectos (49). El proceso de teorización (50). Caracter del proceso de investigación social. El proceso de investigación y los contextos de descubrimiento y justificación (52). Ejercicios (53).

Sección 2. – OPERACIONES BASICAS PRELIMINARES

4. Determinación del problema de investigación 56

Condiciones del problema de investigación (57). Origen del problema de investigación (58). Proceso de determinación del problema de investigación (59). Tipos de problemas de investigación (60). Aspectos de los problemas de investigación (62). Ejercicios (64).

5. Formulación de hipótesis 69

Noción e importancia (69). Naturaleza de las hipótesis: los enunciados (71). Estructura de las hipótesis (74). Tipos de hipótesis (75). Formalización lógica de las hipótesis (79). Determinación de hipótesis (80). Ejercicios (83).

6. Unidades de observación y variables 96

Noción y clases de unidades de observación (96). Noción e importancia de las variables (98). Las unidades de observación y las variables como conceptos y términos. Los conceptos y términos y sus niveles (100). La intensión y la extensión (101). Definición de conceptos (102). Formación de conceptos (104). Tipos de variables (105). Tipos de relaciones entre variables dependientes e independientes (109). Operativización de las variables (110). Las variables empíricas o indicadores (112). Ejercicios (114).

7. El diseño de la investigación social 124

Noción. Importancia y requisitos del diseño (124). La validez y sus aspectos (126). Conceptos conexos con el de validez (127). Aspectos del diseño y operaciones que comprenden (129). Determinación de las variables externas a la investigación de las formas de su control (130). Análisis de las operaciones del segundo aspecto del diseño: desarrollo de la prueba (138). Tipos de diseño y su representación (140). Diseños no experimentales (142). Diseños experimentales (144). Diseños factoriales (148). Ejercicios (150).

8. La evaluación de la investigación 160

Introducción (160). Evaluación del trabajo científico (161). Comprobación (163). Comprobación del proceso de teorización (165). Ejercicios (166).

Sección 3. — MUESTRAS

9. Condiciones y elementos de las muestras 174

Noción (174). Condiciones de las muestras (175). Base de la muestra (176). Unidad de la muestra. Relación entre universo y muestra (178). Ejercicios (181).

10. Elección de la muestra 191

La elección de la muestra y sus requisitos básicos (191). Azar simple y azar sistemático (193). Muestreo estratificado y por cuotas (195). Muestreo de conjuntos y de extracción sucesiva o polietápico. Otros tipos de muestreo (197). Ejercicios (200).

11. El error muestral 206

Noción (206). Fundamento matemático-estadístico del error muestral. Fórmulas del error muestral (208). Estudio de los factores del error muestral (209). Error sistemático (213). Ejercicios (215).

12. Tamaño de las muestras 226

Las fórmulas para hallar el tamaño (226). Afijación de la muestra (227). Ejercicios (228).

Sección 4. — OBSERVACION

13. Observación, medida y escalas 240

Noción general de observación y sus tipos (240). Observación y medida (241). Matemáticas y realidad (243). Medida y escalas (244). Distinción, niveles de medida y clases de escalas (245). Escalas de intervalo (246). Escalas de razón o proporción (247). Ejercicios (249).

14. Observación directa simple 253

Noción y rasgos (253). Sus notas y clases (254). Observación participante (255). Observación precientífica. Observación global (256). Observación descriptiva (257). Observación explicativa (258). Ejercicios (259).

15. Observación experimental 269

Noción (269). Tipos de experimento (271). Operaciones preparatorias del experimento (274). Ejercicios (275).

16. Observación documental 283

Noción y clases (283). Fiabilidad de los documentos escritos (285). Análisis primario de documentos (286). Análisis de contenido. Origen. Noción e importancia (287). Fases del análisis de contenido (289). Elección de categorías (291). Análisis secundario de documentos (291). Ejercicios (294).

17. Cuestionarios 305

Noción y notas de la encuesta (305). El cuestionario y sus clases (306). Tipos de preguntas (307). Reglas para la formulación de preguntas (312). Preparación de cuestionario. Etapas (314). Formulación de hipótesis. Determinación de las variables (315). Planificación del contenido del cuestionario (316). Elaboración formal del cuestionario (318). Juicio crítico de un cuestionario. Los cuestionarios no cumplimentados (319). La no respuesta (320). Validez y seguridad del cuestionario (322). Etica de la investigación por encuesta (323). Ejercicios (324).

18. Entrevista 351

Noción e importancia (351). La interacción social en la entrevista (352). Clases (352). Ventajas de la entrevista. Condiciones formales de la entrevista (356). Preparación de la entrevista. Ejecución de la entrevista (357). Ejercicios (359).

19. Escalas en las ciencias sociales 370

Escalas de actitudes: Fundamento de las escalas de actitudes (371). Condiciones de las escalas de actitudes (372). Tipos de escalas de actitudes (372). Escalas de actitudes arbitrarias (374). Escalas de actitudes experimentales. Otros tipos de escalas de actitudes. (377). Construcción de la escala de Likert (378). Construcción de las escalas de Thurstone (379). Construcción de las escalas de Guttman (380). Cálculo de la seguridad de las escalas sociométricas (382). Validez de una escala (385). Las curvas de opinión pública (387). Ejercicios (389).

Sección 5. — CLASIFICACION

20. Codificación y tabulación 407

Noción y fases de la clasificación (407). Noción de codificación (408). Categorización de preguntas abiertas (409). Formación y lectura de las tablas (410). La explotación del cuestionario (411). La construcción de las tablas (412). Lectura de las tablas (414). Construcción de tipologías (415). Ejercicios (416).

21. Análisis tipológico o de clusters 443

Naturaleza, noción y elementos (443). Tipos (444). Problemas previos (445). Técnicas aglomerativas (446). Técnicas divisivas (447). Ejercicios (447).

Sección 6. — ANALISIS

22. Análisis, interpretación y explicación 458

Noción del análisis, la interpretación y la explicación sociológicos (458). Análisis e interpretación y teorización (459). Clases de análisis (460). Análisis y tipos de datos (461). Análisis e interpretación univariable (462). Análisis e interpretación bivariable (463). Análisis e interpretación multivariable. Nociones de lingüística, lógica, gnoseología, epistemología y ontología de la explicación (464). Tipos de explicación (466). Ejercicios (470).

23. Análisis univariable 476

Resumen de las observaciones sobre una variable (476). Los promedios y los índices de variación. Criterios para la elección del promedio (477). La dispersión relativa y el coeficiente de variación (478). Normalización (479). Proporciones y porcentajes. Ratios o razones, tasas e índices (480). Estandarización. Medida de la desigualdad (481). Las curvas de Lorenz (483). Ejercicios (484).

24. Análisis exploratorio de las distribuciones de frecuencias de Tuckey 490

Objetivos del análisis exploratorio (490). Procedimientos básicos del análisis exploratorio. Los batches (491). Construcción del batch (492). Clases de batches. Características de los batches (493). Representación gráfica (495). Residuales (496). Ejercicios (496).

25. Análisis Bivariable 504

Correlación de variables (504). Coeficientes de correlación (505). Clasificación de los coeficientes de correlación (506). El coeficiente epsilón. El coeficiente Q y Yule, y el coeficiente Phi (508). Coeficiente Lambda (509). El coeficiente Rho de Spearman (515). El coeficiente de correlación r de Pearson (516). El coeficiente Etha (517). El coeficiente O de Wilcoxon (518). Ejercicios (518).

26. Otros coeficientes 535

Coeficientes de correlación biserial y biserial puntual (535). Coeficiente de correlación biserial para rangos o posiciones (536). Prueba de significación. Coeficiente de correlación W de Kendall (537). Coeficiente de consistencia K (539). Ejercicios (540).

27. Test de hipótesis 549

Noción y operaciones que comprenden (549). A. Formulación de supuestos. B. Obtención de la distribución de muestreo (550). C. Selección de nivel de significación (551). D y E. Cálculo del estadístico y decisión (553). Contraste de medias y proporciones en relación a valores dados (555). Comparación de grupos (556). Contraste de los coeficientes de correlación. Análisis de la varianza (559). Operaciones que comprende (561). Test Chi cuadrado (562). Ejercicios (565).

28. Tests no paramétricos 588

Prueba de las secuencias o rachas de Wald-Wollfowitz (589). Prueba de Mac Nemar para la significación de los cambios (590). Prueba de Kolmogorov-Smirnov (591). Prueba de los signos (592). Análisis de varianza no paramétrico: la prueba de Kruskal y Wallis (593). Ejercicios (594).

29. Introducción al análisis multivariable 611

Noción (611). Clasificación (612). Análisis de componentes principales. Análisis de regresión múltiple. Análisis de correlaciones canónicas (615). Análisis factorial. Análisis discriminante (616). Análisis de varianza múltiple. Análisis de covarianza. Análisis de senderos. Análisis de sistemas (617). Análisis Log-lineal. Análisis de Cluster. Análisis de correspondencias. Escalonamiento multidimensional (618). El análisis multivariable y los modelos matemáticos. Operaciones que comprenden los modelos (619). Ejemplo de construcción de un modelo sociológico (621). Ejercicios (622).

30. Análisis causal 630

Métodos experimentales y correlacionales. Relación de causalidad (630). Procedimiento de análisis de los modelos multiecuacionales causales en las ciencias sociales (632). Análisis multivariable de Lazarsfeld (633). Procedimiento (635). Origen y presupuestos del "Path analysis" (638). Elementos del análisis de senderos (639). El modelo de senderos y las ecuaciones estructurales (640). Coeficiente de Wright. Efectos total, directo e indirecto (641). Interpretación. Ejemplo numérico de análisis de senderos (642). Coeficientes d (644). Ejercicios (645).

31. Análisis del cambio social 655

El análisis estadístico de las series temporales (656). Fundamentos teóricos del modelo (657). Representación gráfica del modelo y especificación de sus ecuaciones matemáticas (658). Obtención de los datos empíricos sobre las variables del modelo. Tablas de rotación (659). Tablas de rotación de la movilidad social (660). Análisis descriptivo de las tablas de movilidad social (662). Tabla de movilidad perfecta y sus índices (665). Tabla de dieciseis casillas (667). Análisis cuantitativo y cualitativo de las tablas (669). Análisis cualitativo. Análisis de la causalidad (670). Ejercicios (671).

32. Análisis sociométrico 683

Nociones básicas de la sociometría (683). El test sociométrico (684). El análisis sociométrico (687). Valores sociométricos. Indices sociométricos (688). El sociograma (692). Análisis individual y del grupo en su conjunto (693). Ejercicios (694).

Bibliografía 705

INDICE DE TABLAS

De números aleatorios 202
De error muestral para un nivel de seguridad de 3 sigmas 219
De error muestral para un nivel de seguridad de 2 sigmas 220
De tamaño de las muestras para poblaciones infinitas o muy numerosas e intervalo de confianza de 3 sigmas 231
De tamaño de las muestras para poblaciones infinitas o muy numerosas e intervalo de confianza de 2 sigmas 232
De tamaño de las muestras para poblaciones finitas y 3 sigmas de intervalo de confianza 233
De tamaño de las muestras para poblaciones finitas y 2 sigmas de intervalo de confianza 234
De las áreas de la curva normal 396
De variación de x^2 567
De valores significativos de G para contrastar H_0 574
De valores significativos de R para contrastar H_0 577
De valores de F para el análisis de varianza 578
De valores de t para contrastar H_0 583
De valores críticos de r_s, el coeficiente de correlación de rangos de Spearman 603
De probabilidad asociadas con valores tan grandes como los valores observados de S en el coeficiente de correlación de rango de Kendal . . 604
De valores críticos de S en el coeficiente de concordancia de Kendal . . 605
De valores críticos de r en la prueba de rachas 606
De valores críticos de D en la prueba de una muestra Kolmogorov-Smirnov 607
De valores críticos de K_D en la prueba de dos muestras de Kolmogorov-Smirnov (muestras pequeñas) 608
De valores críticos de D en la prueba de dos muestra de Kolmogorov-Smirnov (Muestras grandes: prueba de dos colas) 609
De probabilidades asociadas con valores tan pequeños como los valores observados de x en la prueba binomial 610

PRESENTACION DE LA UNDECIMA EDICION

Publicada esta obra por primera vez en el año 1975, las sucesivas ediciones que ha tenido, aparecidas con modificaciones a veces importantes, especialmente en el caso de la segunda y sexta edición, prueban su aceptación por los profesores, estudiantes e interesados en las técnicas de investigación social.

Dado el tiempo transcurrido a partir de la sexta edición, para preparar esta undécima, se ha procedido a una nueva revisión de la materia y se ha llegado a la conclusión de que el libro conserva su validez como visión general y básica de las técnicas de investigación social.

Sin duda, la innovación más importante ocurrida en este período, es la generalización de la aplicación de los programas informáticos para realizar mediante ordenadores las operaciones estadísticas que comporta el análisis de los datos obtenidos en las investigaciones sociales. Se trata pues, de una materia que al referirse al manejo de los ordenadores, forma parte de la informática y no pertenece propiamente, por tanto, a las técnicas de investigación social. Obras específicas que se ocupan de la aplicación del paquete informático SPSS/PC±, que es el más usado en el análisis estadístico de los datos obtenidos en las investigaciones sociales, referenciadas en la bibliografía, son las de J. J. Sánchez Carrión, 1944 y J. Etxeberría y otros, 1990.

No obstante, en esta nueva edición, se ha ampliado el contenido de algunos de los puntos tratados y se han añadido en la bibliografía las refencias de obras aparecidas últimamente.

Sección 1

FUNDAMENTOS

1. NOCION Y FUNDAMENTO DE LA ASIGNATURA

Noción y fundamento.– En el título de la disciplina que nos va a ocupar en este libro, se pueden distinguir dos partes, unidas por el genitivo de: técnicas e investigación social. Examinémoslas por separado.

La palabra técnica etimológicamente tiene dos sentidos: el de arte y el de forma de actuación. En ambos, arte práctico en general o modo de actuación particular, la técnica se contrapone en griego, y también en las ciencias modernas, a la *episteme*: la ciencia, en cuanto ésta es conocimiento o saber teórico y como tal no inmediatamente práctico.

En los lenguajes modernos, y entre ellos el castellano, se ha producido una separación de los dos sentidos indicados. Las actividades prácticas y artísticas se designan con la palabra arte, mientras que el término técnica ha quedado reservado para las formas o procedimientos de actuación práctica.

El Diccionario de la Real Academia Española, define la palabra técnica como "el conjunto de procedimientos o recursos de que se sirve una ciencia o un arte". En esta definición aparece la amplitud del significado de técnica y su relación con la ciencia.

Técnica es una palabra que tiene una significación muy amplia. Comprende los procedimientos o formas de realizar las distintas actividades humanas, incluso intelectuales, y el modo de utilización de los instrumentos y máquinas que utiliza el hombre, así como las maneras de preparar esos instrumentos.

De las dos áreas que comprende el conocimiento, las artes, entendidas en sentido amplio, y las ciencias, entendidas como actividad orientada a ampliar nuestro conocimiento del mundo que nos rodea, las técnicas de investigación social se inscriben en el área científica, y son técnicas no artísticas sino científicas. Ello es así porque atañen a la Sociología y a las demás ciencias de la sociedad.

El primer fundamento de las técnicas en nuestro caso es, pues, la ciencia en cuanto deben a ella su razón de ser y a ella se orientan.

LA CIENCIA

La noción de ciencia.– La palabra ciencia se deriva etimológicamente en las lenguas modernas del vocablo latino *scientia*. En el latín ciencia tiene un sentido muy amplio y significa: conocimiento, práctica, doctrina, erudición. Esta concepción latina de ciencia concuerda con el significado del origen de su raiz, el verbo latino *scio* que se deriva a su vez del griego *"isemi"*. Este verbo griego equivale también a saber, en toda la extensión de la palabra: conocer, tener noticia de, estar informado.

Por tanto, ciencia, en su acepción original y más general, equivale a toda clase de saber. Sin embargo, históricamente, ciencia vino a significar, de modo más preciso, un conjunto de conocimiento sistematizado sobre una materia. Esta noción coincide con lo que hoy entendemos por disciplina. Y así, en este sentido, se puede decir que en la Edad Media las ciencias o disciplinas por antonomasia fueron la filosofía y la teología.

Con el desarrollo del saber experimental y la insistencia en el método inductivo, el concepto de ciencia ha quedado reservado modernamente para el conocimiento teórico, inductivo y sistemático sobre la realidad, derivado de la observación y experimentación metódicas.

Definición y elementos.– La ciencia se puede definir, en sentido estricto, como un conjunto de conocimientos sobre la realidad observable, obtenidos mediante el método científico.

Según esta definición son tres los elementos esenciales de la ciencia, que configuran su naturaleza: Un contenido, un campo de actuación y un procedimiento o forma de actuar.

La ciencia en cuanto a su contenido está constituida exclusivamente por un conjunto de conocimientos sobre la realidad, en forma de términos y de enunciados. Las ideas de este conjunto se hallan interrelacionadas entre sí y forman lo que se llama la teoría. De éstas nos ocuparemos después más detenidamente.

Es necesario darse cuenta de que la ciencia aunque se refiere a la realidad empírica, no está formada por hechos sino en todo caso por ideas. El hombre solo puede captar la realidad conceptualmente.

El campo de actuación propio y único de la ciencia es la realidad observable, la realidad de este mundo en que vivimos. Lo no empírico, digamos lo trascendente, cae fuera del campo de la ciencia en sentido estricto.

Por último, la ciencia utiliza como procedimiento o forma de actuación en la formación del conjunto de conocimientos que la integran, el método científico, que es el que la tipifica, como se ve a continuación.

Lo específico de la ciencia.– Lo más característico y específico de la ciencia, aquello que la constituye como tal y la distingue de los demás tipos de conocimiento es el método científico. La ciencia no se distingue de otros tipos de conocimiento por los objetos sobre los que versa.

Se puede tratar intelectualmente los fenómenos físico-químicos y los sociales de múltiples maneras: por ejemplo, artística, filosófica, literaria, etc., pero sólo puede decirse que se estudian científicamente si se utiliza el método científico. Su empleo respecto a la realidad social es lo que hace, como en cualquier otro campo de la realidad o del conocimiento, que las disciplinas sociales sean propiamente ciencias.

Tampoco es distinto esencialmente el contenido de las ciencias del de los demás tipos de conocimiento, este contenido es, en todos los casos y en último término, ideas.

La ciencia, pues, no es otra cosa que un conjunto de ideas obtenidas por la aplicación del método científico, y, por tanto, es el resultado de este método.

Por ello, un científico es, ante todo, no el que tiene muchos conocimientos sobre una materia determinada, lo que es importante y valioso, sino quien sabe utilizar correcta y eficazmente el método científico en su campo.

De aquí la importancia del método científico. A él hay que atribuir en gran parte, sin duda, el extraordinario desarrollo de la ciencia. Se puede decir que la ha posibilitado.

No está demostrado que los hombres a partir del siglo XVII sean más inteligentes que los anteriores. Sin embargo, han hecho avanzar grandemente la ciencia porque abandonaron el método escolástico e iniciaron el método científico.

A pesar de ello el método científico no lo es todo. Es preciso también inteligencia, imaginación e intuición.

Clases de ciencias.– Por su objeto las ciencias se pueden dividir en: ciencias físico-naturales, ciencias humanas y ciencias sociales.

En las primeras la realidad observable a que se refieren está constituida por toda la naturaleza, universo sensible o mundo, excepto el hombre y su sociedad, que son respectivamente el objeto y la realidad observable peculiar de las ciencias humanas y sociales.

Otra clasificación importante de las ciencias es la que las divide en nomotéticas e idiográficas.

Las primeras son las que estudian el aspecto regular y repetitivo de los fenómenos, con el fin de hallar leyes (nomos) de aquí su nombre.

A ellas pertenecen las ciencias sociales por lo general: la Sociología, la Economía, la Política, etc.

En cambio, las ciencias idiográficas versan principalmente sobre cosas o fenómenos únicos, singulares, no múltiples y repetidos como las nomotéticas. El ejemplo que se suele señalar como típico de estas ciencias es la historia.

También se distinguen las ciencias sincrónicas de las diacrónicas. Las primeras, como señala Galtung, tratan de fenómenos que tienen lugar en cualquier sitio en el espacio, pero dentro de un intervalo relativamente corto de tiempo, mientras que las segundas tratan de fenómenos históricos que tienen lugar en cualquier punto del tiempo pero dentro de una región limitada del espacio.

Objetivos de la ciencia.– Los objetivos fundamentales de la ciencia, o del conocimiento científico, en relación a su campo de actuación, la realidad de este mundo, son cuatro: analizar, explicar, prever o predecir y actuar.

El primer objetivo de la ciencia es saber cómo es la realidad, qué elementos la forman y cuáles son sus características.

Después de conocer cómo es la realidad, el segundo objetivo de la ciencia es explicarla, llegar a establecer cómo se relacionan sus distintas partes o elementos, porqué es como es la realidad.

Estos son los objetivos básicos y principales de la ciencia. Su consecución capacita a la ciencia para alcanzar los otros dos objetivos indicados, que por ello son derivados o aplicados, la predicción y la actuación.

Por una parte, si la ciencia logra saber cómo es un sector de la realidad y los factores que la explican, entonces está en condiciones de prever los acontecimientos que tendrán lugar en dicho sector de la realidad.

Por otra parte, el mismo conocimiento del cómo y por qué de un sector de la realidad, faculta también para actuar, da poder para transformar esa realidad e influir en ella, en mayor o menor grado.

La ciencia de hecho, en nuestros días, ha concedido un poder inmenso al hombre. Este poder es peligroso, ya que puede ser utilizado para el bien o para el mal.

De aquí que B. Russel diga (1969, p. 219) que "para que la civilización científica sea una buena civilización es necesario que el aumento

de conocimientos vaya acompañado de sabiduría. Entiendo por sabiduría, dice, una concepción justa de los fines de la vida. Esto es algo que la ciencia por sí misma no proporciona".

LOS METODOS DE INVESTIGACION

El método, como forma de realizar la actividad humana se divide en dos grandes clases, según que dicha actividad sea de carácter humano interno, métodos de pensamiento, o de carácter externo, métodos de acción propiamente dicha.

Esta última clase, los métodos de acción, se pueden subdividir a su vez en,

- técnicas o métodos de actuación dirigidas a manipular y transformar la realidad y
- métodos de investigación, que son formas de actuación humana orientadas al conocimiento de la realidad observable, del mundo que nos rodea.

Aquí corresponde referirse ahora a los métodos de investigación, dado que los métodos de pensamiento y las técnicas o métodos de trabajo y producción caen fuera del ámbito de esta obra.

Un método de investigación completo debe comprender no sólo un contenido determinado, o la especificación de una serie de fases o etapas a seguir para lograr el conocimiento pretendido, sino también una base racional, constituida por:

a) Los presupuestos filosóficos del método de investigación;

b) el enfoque adoptado por el método,

c) los principios racionales que orienten y justifiquen las actuaciones que el método de investigación suponga, y

d) técnicas específicas para llevar a efecto, según las circunstancias del caso, las fases y operaciones del método de investigación.

Que los elementos que se puedan distinguir en el método de investigación sean todos estos, no quiere decir que todos los que se llaman o presentan como tales los posean.

De todos estos métodos, el que se pueda considerar como método de investigación por excelencia, es el método de investigación científico, que

suele llamarse, no con mucha propiedad, método científico sin más. Es así, porque su razón de ser y su objetivo es el conocimiento de la realidad, comprende todos los elementos indicados y se presenta acompañado o dispone de técnicas apropiadas y las utiliza.

Hay que repetir que el método de investigación científico, tiene, como método, un carácter instrumental, que lo hace susceptible de emplear por todas las ciencias, cualquiera que sea el campo de la realidad a que se refieran. Por ello, aunque tuviera su origen en las ciencias físicas y naturales, es independiente de ellas y no está ligado necesariamente a las mismas.

EL METODO DE INVESTIGACION CIENTIFICO. SU NOCION Y RASGOS

El método de investigación científico, como tal método, es un procedimiento de actuación general seguido en el conocimiento científico. Ahora bien el procedimiento científico, como todo procedimiento, se concreta en conjunto de trámites, fases o etapas. Por ello parece que la mejor manera de expresar en qué consiste es describir los trámites o actuaciones que comprende.

Enfocado de este modo, el método de investigación científico consiste en formular cuestiones sobre la realidad del mundo y la humana, basándose en las observaciones de la realidad y en las teorías ya existentes, en anticipar soluciones a estos problemas y en contrastarlas con la misma realidad, mediante la observación de los hechos, su clasificación y su análisis.

Según Isaac Asimov (1979), el método científico, en su versión ideal, consiste en:

1. Detectar la existencia de un problema.
2. Separar luego y desechar los aspectos no esenciales.
3. Reunir todos los datos posibles que incidan sobre el problema, mediante la observación simple y experimental.
4. Elaborar una generalización provisional que los describa de la manera más simple posible: un enunciado breve o una formulación matemática. Esto es una hipótesis.
5. Con la hipótesis se pueden predecir los resultados de experimentos no realizados aún y ver con ellos si la hipótesis es válida.

6. Si los experimentos funcionan, la hipótesis sale reforzada y puede convertirse en una teoría o una ley natural.

En esta descripción del método de investigación científico no sólo se comprende sus distintas fases, sino que de ella se deriva su carácter complejo.

El método científico en el estado actual de las ciencias es en primer lugar un método **a) teórico** en su origen y en su fin. Con ello se quiere decir que su punto de partida es, en general, una teoría previa o un conjunto racional y sistemático de ideas sobre la realidad de que se trate. Esta teoría debe ser normalmente la fuente de los problemas que formula el método científico.

Es también su fin, porque de los resultados de la observación e inducción empírica realizadas, se deben deducir nuevos principios que reformen, completen o confirmen las teorías iniciales. Además, también es necesaria la teoría para observar la realidad. Los hechos de por sí son mudos y nada dicen si no se sabe interpretarlos y se va a ellos con ideas y enfoques previos.

b) problemático-hipotético, en cuanto se basa en la formulación de problemas, cuestiones o interrogantes sobre la realidad y en adelantar conjeturas o soluciones probables a dichas cuestiones.

Es **c) empírico,** en el sentido de que su fuente de información y de respuesta a los problemas que se plantea, es la experiencia. Que la fuente de información y de respuesta del método científico es la experiencia, quiere decir que la ciencia, a efectos de la prueba en que consiste, toma sus datos y funda sus conclusiones en la observación ordenada y sistemática de la realidad. En esto se diferencia de otras formas de conocimiento que basan también sus pruebas en la autoridad, la tradición y la revelación.

Es **d)** a la vez, **inductivo y deductivo**. Es inductivo en cuanto procede mediante la clasificación sistemática de los datos obtenidos mediante la observación, con el fin de determinar las uniformidades o regularidades que presentan.

La ciencia, aunque se base en la inducción sistemática en mayor medida que otros tipos de conocimiento, utiliza, asimismo, en gran medida la deducción.

Esta, como se sabe, consiste en la derivación de conceptos o enunciados, no de la observación de la realidad, como la inducción, sino de otros conceptos o enunciados establecidos anteriormente.

La inducción y la deducción en la ciencia no se opone entre sí, sino que la deducción está íntimamente unida en ella a la inducción.

La inducción sólo da lugar inmediatamente a datos sobre la realidad. Pero el relacionar estos datos, establecer conceptos y enunciados con base en ellos y sacar conclusiones de todo género es en gran parte obra deductiva.

e) crítico. Con ello se quiere decir:

1. Que debe someter constantemente a crítica o examen y juicio, todas sus fases, operaciones y resultados, o lo que es lo mismo, a contraste y verificación.
2. Que en ningún caso los logros del método científico son definitivos y que siempre están sujetos a la revisión, que se puede derivar de nuevos descubrimientos y puntos de vista científicos.

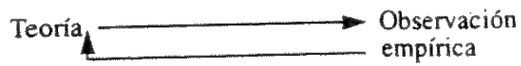
El científico, escribe E. B. Wilson (1952, 21) "rechaza la autoridad como el fundamento último de la verdad. Aunque se ve obligado por necesidad práctica a usar hechos y proposiciones establecidas por otros investigadores, se reserva la decisión sobre si ellos merecen confianza, si sus métodos son buenos y si, en un caso particular, los hechos alegados son creíbles. El, además, considera un privilegio y a veces un deber, repetir y comprobar el trabajo de otros siempre que estime que esto es deseable".

f) circular. Es así porque como señala el profesor Salustiano del Campo (1969, p. 267) en él "los principios se prueban mediante datos empíricos y éstos se analizan o interpretan sobre la base de aquellos".

Existe, pues, una interacción continua en el método científico entre la experiencia y la teoría: con base en la experiencia se establece, completa y reforma la teoría, y con base en la teoría se capta y explica la realidad.

El modelo del método científico, y por tanto de la ciencia, no es simplemente lineal del tipo siguiente:

sino que se ajusta al siguiente esquema:



g) analítico-sintético. Es decir, estudia la realidad distinguiendo y separando unos de otros sus elementos más simples, pero no se queda aquí sino que procura luego unir y recomponer los elementos separados obteniendo una nueva visión global del conjunto y de las relaciones estructurales entre sus elementos.

h) selectivo, en un doble sentido. Primero entre la multiplicidad de aspectos de los fenómenos, debe concentrar su observación en los más relevantes, y segundo, entre la masa de datos recogidos debe procurar detectar en el análisis los más significativos por tener un influjo predominante.

i) debe atenerse normalmente a las reglas metodológicas formales, pero al mismo tiempo **debe fomentar la intuición y la imaginación aun en el caso** de que no se atenga con ello estrictamente a dichas reglas, e incluso a las teorías admitidas, como señala P.K. Feyerabend en su obra "Contra el método", Ed. Ariel.

En resumen, los caracteres del método científico son, según lo anterior, teórico, problemático-hipotético, empírico, inductivo, deductivo, crítico, circular, analítico-sintético, selectivo y abierto a la imaginación.

Sin embargo, ello no quiere decir que todos esos caracteres sean exclusivos del método científico. De hecho, otras disciplinas no empirico-científicas, por ejemplo, la filosofía y la teología, emplean incluso la inducción y la experiencia.

Frente a ellas, lo que caracteriza al método científico es: 1, la ordenación y disposición conjunta de dichos caracteres; 2, el tener un campo único de acción, que es la realidad observable; 3, el admitir, en último término, como fuente de información y de prueba sobre esta realidad exclusivamente la experiencia.

EL METODO CIENTIFICO EN LAS CIENCIAS SOCIALES

La vida social, en sus diversas manifestaciones, es el objeto de las ciencias sociales. Esta vida constituye una parte del mundo o de la realidad observable en la que vive inmerso el hombre. Por ello, como parte de esta realidad, puede ser estudiada científicamente, aplicando a ella el método científico.

En efecto, es innegable que se pueden formular cuestiones o problemas referentes a la vida social, anticipar soluciones a ellos y tratar de verificarlas mediante la observación sistemática de los fenómenos

sociales, la clasificación de estos fenómenos y su análisis y explicación. Precisamente la posibilidad de aplicación del método científico a los fenómenos sociales es lo que hace que las ciencias sociales lo sean verdaderamente.

No obstante lo anterior hay que tener en cuenta que el método científico en las ciencias sociales presenta algunas particularidades especiales en relación con las ciencias hoy típicas, las naturales, debido a las características peculiares de su objeto de estudio, la sociedad.

La sociedad es una realidad constituida en último término por relaciones sociales, que aunque se materialicen de formas diversas, son en su raiz inmateriales. Esta sociedad en sus manifestaciones materiales es muy diversa y compleja, sujeta a constante evolución y cambio. De ella forman parte importante elementos inmateriales: principios, creencias, reglas morales, normas jurídicas, etc, pertenecientes al mundo de los valores, de las ideas y del deber ser. Está sujeta, además, a la incertidumbre e indeterminación propia de la libertad del hombre, elemento consitutivo de la sociedad.

Todo ello se traduce en la dificultad, por una parte, de medida de los fenómenos sociales y, por otra, de establecimiento de regularidades y generalizaciones respecto a ellos. A esto se une la ausencia de instrumentos de observación tan potentes como en las ciencias físicas, y la dificultad de aplicar las técnicas de observación experimentales. La consecuencia es que, en general, no se puede emplear en las ciencias sociales el método científico de modo tan riguroso como en las naturales, ni es posible, hasta ahora, llegar en ellas a resultados tan exactos, exhaustivos, constantes y generales como en ellas.

Es obvio que las dificultades de aplicación del método científico según los distintos campos, no impide que sea uno y el mismo en todos ellos.

Esta conclusión plantea la cuestión del carácter que se debe atribuir a los llamados métodos, por ejemplo, funcional, estructuralista y dialéctico, etc, de que tanto se habla en las ciencias sociales.

En relación a este problema hay que distinguir, a mi juicio, en el método, su contenido y sus presupuestos filosóficos. Su contenido está formado por el conjunto de reglas y procedimientos de actuación seguidos en el trabajo científico.

Sus presupuestos filosóficos son aquellos principios racionales en que se basa todo método, por ejemplo, referentes a la existencia de la realidad, a la posibilidad de su conocimiento y a la concepción de la esencia y caracteres de esa realidad.

Admitida esta distinción, el método dialéctico y el funcionalista se refieren, a mi juicio, a los presupuestos filosóficos del método científico. En concreto, representan formas distintas de enfocar la realidad.

Y así, mientras el método funcional-estructuralista concibe o mira la realidad, ante todo estáticamente, el método dialéctico la ve dinámica y conflictivamente.

En mi opinión, sin negar la validez de estos métodos y sin defender una posición positivista, estimando como tal la que defiende el modelo de las ciencias naturales y sus métodos propios como únicamente válidos para todas las ciencias, incluidas las humanas y sociales, la realidad es que dichos métodos y otros similares no constituyen una alternativa válida que pueda sustituír al método científico en las ciencias sociales, en cuanto no han desarrollado hasta ahora técnicas específicas propias, distintas de las usadas por el método científico.

Con esta cuestión guarda relación la polémica sostenida por Adorno y sus compañeros de la Escuela de Frankfurt con Popper y sus seguidores, como puede verse en Adorno y otros (1973) y Beltrán Villalba (1979).

CLASES DE METODOS EN LAS CIENCIAS SOCIALES

Aunque el método puede presentar diversas modalidades, especialmente en las ciencias sociales, se distingue ante todo según se centre, por ejemplo; en la observación de muchos casos particulares o en el estudio a fondo y globalmente, cualquiera que sea su amplitud, de uno sólo o unos pocos casos individuales.

En el primer caso, se tiene el método cuantitativo predominantemente inductivo, que busca determinar las características externas generales de una población basándose en la observación de muchos casos individuales de la misma.

En el segundo caso se trata de los métodos científicos cualitativos o si se quiere fenomenológicos que pretenden comprender, lo más profundamente posible, una entidad, fenómeno vital o situación determinada.

Las diferencias entre ambas modalidades del método científico consisten en que 1° de las tres etapas fundamentales del método científico, observación, clasificación y análisis, todas ellas tienen un acusado carácter numérico en los métodos cuantitativos, mientras que lo que se pretende en los cualitativos es informar de las "observaciones en lenguaje natural" (Schwartz, 1984:23), y 2° en los métodos primeros, la interpretación y la explicación tienen un carácter objetivo más marcado, mientras que en los segundos presentan un mayor carácter subjetivo, pues se basan en la comprensión íntima de la realidad y la captación de su sentido, y, por tanto, en la capacidad de intuición personal de los investigadores.

Asimismo, la prueba empírica tiene mayor vigencia en los métodos inductivos, mientras que en los comprensivos tal prueba se centra más bien en el consenso de la comunidad científica, si bien tal consenso no deja de tener importancia también en los primeros, como lo ha puesto de manifiesto T.S. Kuhn (1975).

En opinión de H. Schwartz (1984:22)

> "Los métodos cualitativos, que utilizan el lenguaje natural, son mejores para obtener acceso al mundo de la vida de otras personas en breve tiempo. Por otra parte, los métodos cuantitativos son mejores para conducir una ciencia positiva, esto es, permiten una recolección de datos clara, rigurosa y confiable y permiten someter a prueba hipótesis empíricas en una forma lógicamente consistente."

METODO CIENTIFICO Y TECNICAS CIENTIFICAS

Como ya se ha indicado, el método científico es el procedimiento de actuación general seguido en el conocimiento científico. Igualmente, de lo expuesto hasta ahora se puede deducir que el método científico comprende las fases fundamentales, indicadas ya, de actuación en la actividad científica y las normas básicas a seguir en ellas.

También se puede deducir de la exposición que hice sobre la noción de técnica, que técnicas científicas son los procedimientos concretos de actuación, operativos, que cabe utilizar dentro de las ciencias, para llevar a efecto las distintas etapas del método científico.

La relación existente, pues, entre método científico y técnicas científicas parece clara. Su naturaleza es la misma. Ambos son procedimientos, formas de actuación científica. Su diferencia consiste en su amplitud. El método es el procedimiento general de conocimiento científico y es común, en lo fundamental, a todas las ciencias. Las técnicas por el contrario, son procedimientos de actuación concretos y particulares, relacionadas con las distintas fases del método científico.

Aunque puede haber técnicas comunes a todas o a varias ciencias; las técnicas lógicas y las matemáticas sobre todo, sin embargo, en general cada ciencia o grupo de ciencias tiene sus técnicas específicas.

2. LA INVESTIGACION SOCIAL

Noción y caracteres.— En la lección anterior hemos estudiado la noción del primer elemento del título de la disciplina, las técnicas y los fundamentos de éstas, la ciencia y el método científico.

Ahora nos toca examinar el segundo elemento del nombre de la disciplina, la investigación social.

Genéricamente, la investigación es una actividad humana, orientada a descubrir algo desconocido. Tiene su origen en la curiosidad innata de los hombres, es decir, en su deseo de conocer cómo y por qué son las cosas y cuáles son sus razones y motivos. Responde también a la indigencia natural del hombre, al que la naturaleza no le ha dado todo resuelto como a los animales, lo que le obliga a investigar y buscar solución a sus problemas, dificultades y necesidades.

Según esto, toda averiguación sobre algo no conocido y toda búsqueda de solución a algún problema es investigación, pero sólo será investigación científica si actúa de este modo, es decir, según un método científico, y sólo lo será social si su campo de investigación es la sociedad.

De acuerdo con lo anterior, se puede definir la investigación científico social como el proceso de aplicación del método y técnicas científicos a situaciones y problemas concretos en el área de la realidad social para buscar respuesta a ellos y obtener nuevos conocimientos.

En ésta definición de investigación se pueden distinguir las siguientes notas:

1. La investigación en primer lugar es un proceso formado como tal por un conjunto de fases de actuación sucesivas, orientadas en este caso a descubrir la verdad en el campo social.

   Al ser un proceso, la investigación se distingue por ello del método científico, que no es un conjunto de actividades, sino un conjunto de normas y reglas genéricas de actuación científica. Por el mismo motivo, la investigación se distingue de las técnicas en que, como he indicado ya, son también procedimientos, aunque más concretos que el método.

2. En segundo lugar, es fundamental que la investigación científico social tenga como finalidad hallar respuesta a problemas desconocidos y ampliar el ámbito de nuestros conocimientos en el área social.
3. En tercer lugar, la investigación científico social exige necesariamente, la aplicación lo más rigurosa posible del método y las técnicas científicas al campo social.
4. En cuarto lugar, la investigación debe referirse a problemas concretos, es decir, lo más precisos y específicos que sea posible, y reales o sea, no especulativos, sino referentes a la realidad social.

PECULIARIDADES DE LA INVESTIGACION SOCIAL Y SUS PROBLEMAS

Las peculiaridades de la investigación social se pueden sintetizar, abundando en lo expuesto al tratar del método científico en las ciencias sociales, como sigue: (Se puede consultar, además H.M. Blalock, Jr. 1979)

En primer lugar, cabe señalar el carácter cualitativo de muchos de los fenómenos y de los aspectos de la realidad social. Este caracter comporta dos dificultades principales, a mi juicio, en la investigación social.

1ª Lo externo, lo material digamos, de las acciones y los acontecimientos sociales, lo que aparece al observador, no es siempre lo más importante en los mismos, desde el punto de vista científico. Por debajo de ellos laten frecuentemente intereses, ideologías y actitudes de los sujetos sociales que intervienen en ellos. Estos aspectos son a veces los verdaderamente importantes para la investigación social, pero por su carácter íntimo e inmaterial, no se pueden observar ni tampoco medir y cuantificar directamente.

2ª Por otra parte, la investigación social versa frecuentemente sobre ideas morales que no tienen una realidad objetiva externa y que sólo se pueden investigar por procedimientos indirectos.

Conceptos tales como la solidaridad, la alienación, el autoritarismo, etcétera, sólo se pueden observar empíricamente mediante la búsqueda de indicadores que sean expresión de ellos.

Consecuencia de todo ello es la dificultad de medida de los fenómenos sociales en general. Estos son incuantificables a veces o sólo se pueden cuantificar en términos muy poco precisos por lo general.

La dificultad de medida aleja las ciencias sociales de la exactitud que constituye el ideal de la ciencia, y es causa de que los coeficientes de error con que se trabaja en ellas sean mucho más elevados que en las ciencias físicas.

Por ello, gran parte del contenido actual de las ciencias sociales pertenece, en un grado mayor o menor, al reino de la especulación, en el que el acuerdo es muy escaso, pues cada uno tiene su opinión.

El segundo lugar, está la multiplicidad de factores de los fenómenos sociales. Implica que en todos ellos intervienen un gran número de variables, generalmente en interacción mutua y no simplemente causas o efectos solamente, e incluso algunas desconocidas. Entre ellas hay que contar la misma actuación investigadora que constituye también una variable que influye en la situación que investiga.

En tercer lugar, se tiene el acusado grado de variabilidad de los fenómenos sociales en el espacio y en el tiempo. De aquí que el establecimiento de regularidades y generalizaciones que es una de las tareas fundamentales de la ciencia, haya de ser mucho más prudente y limitado que en otras ciencias.

En cuarto lugar cabe destacar la ausencia hasta ahora en las ciencias sociales de instrumentos de observación, potentes y precisos. Por ejemplo, no existe aún en estas ciencias un instrumento de observación que se pueda comparar, ni aún de lejos, con el microscopio en las ciencias biológicas, físicas y químicas, el telescopio en la astronomía y los rayos X en la medicina.

Por otra parte, en las ciencias sociales sólo se puede utilizar muy imperfectamente y con muchas limitaciones el experimento, que es la técnica científica de observación más potente.

En quinto lugar, hay que destacar otra característica muy peculiar de la investigación social, que aumenta la dificultad del estudio científico de los fenómenos sociales.

Consiste en la innegable influencia que tienen en el mismo objeto de investigación, la sociedad, la difusión en ella de los resultados de la investigación.

Como escribe S. Andreski (1972, p. 22), "podemos imaginar el triste aprieto en que se vería el científico natural si los objetos de la investigación tuvieran la costumbre de reaccionar ante lo que se dice acerca de ellos: si las sustancia pudieran leer u oir lo que el químico escribe o dice acerca de ellas y pudieran saltar de sus recipientes y quemarlo cuando las desagradara lo que ven en la pizarra o en el cuaderno de notas".

Por último, no debe ser omitido el hecho de que el investigador forma parte de la sociedad que investiga y participa de sus valores, ideologías y creencias. Ello hace que nunca pueda ser totalmente independiente y neutral respecto a la sociedad investigada.

Esta última peculiaridad se puede decir que se identifica con el problema de los juicios de valor en la investigación social. Este problema ha sido objeto de discusión especial por los científicos sociales, lo que justifica su examen por separado.

Hay que advertir que con todo lo expuesto no se pretende insinuar que la investigación científica social sea imposible, sino únicamente avisar sobre sus dificultades. Estas, por otra parte, también existen en todas las ciencias, aunque en las sociales sean más acusadas, y exijan por tanto, quizás, en el investigador una mayor capacidad de observación de intuición y de raciocinio.

La investigación social es posible, como lo demuestran muchos conocimientos adquiridos, descubrimientos realizados y aplicaciones prácticas de las ciencias sociales. Y no sólo es posible, sino que también es muy necesaria porque urge reducir distancias entre el extraordinario adelanto de las ciencias naturales y el incipiente desarrollo de las ciencias sociales y de sus técnicas, que parece ser una de las causas más importantes del grave desequilibrio de las sociedades modernas.

Es también necesaria para lograr, no ventajas materiales como las que se derivan de las ciencias naturales por lo general, sino la construcción de una sociedad más justa, más humana, que permita una vida feliz y un desarrollo pleno de todos los hombres.

EL PROBLEMA DE LOS VALORES EN LA INVESTIGACION SOCIAL

El problema de la influencia de las ideologías, de los llamados juicios de valor, en la actividad y en las formulaciones científicas en el campo social, fue planteado en las ciencias sociales ante todo por Max Weber. En sus obras defendió lo que él llamaba la neutralidad valorativa, es decir, una posición del científico social de libertad e independencia frente a los juicios de valor.

Su criterio motivó una fuerte polémica aún no acabada y que todavía sigue sin solución definitiva.

Muchos autores opinaron que es imposible una investigación social ajena a la influencia de los valores. En esta línea hay que incluir el pensamiento marxista que, tanto insiste en la importancia de las ideologías en la actividad y las realizaciones humanas.

En mi opinión, las dos posiciones tienen su parte de verdad, pero ambas son rechazables si se toman exclusivamente.

Por una parte, parece indudable que las actividades y concepciones, incluso científicas, están en relación y, sin duda, influidas, por su paradigma, o concepción general del mundo, de la vida y de la ciencia vigente en un momento histórico.

Y así se puede afirmar que diferentes paradigmas darán o pueden dar lugar a lecturas más o menos distintas de la misma realidad.

T.S. Kúhn en su conocida obra, "La estructura de las revoluciones científicas" sostiene, incluso, que toda la ciencia de una época está montada sobre un paradigma determinado, y es cuando este paradigma se demuestra inadecuado cuando se producen las revoluciones científicas para sustituir el antiguo por un nuevo paradigma.

Pero, por otra parte, está claro que el carácter empírico y positivo de las ciencias sociales, dada su conexión más estrecha con los valores que las ciencias naturales, exige que la actividad investigadora de los científicos sociales sea lo menos posible mediatizada por sus juicios de valor particulares.

Esta exigencia no sólo mira a salvaguardar el trabajo científico de la influencia de los valores e ideologías, sino que también restringe y limita el papel de la ciencia a su propio campo, el de la experiencia. De tal modo es así que si formula juicios de valor o saca conclusiones valorativas, una vez obtenidos los resultados científicos, traspasa su propio campo y entra en el de la filosofía, en el de la ética o en el de la política.

La objetividad de las ciencias sociales, entendidas de esta manera, no supone una concepción aséptica de éstas ni es contraria a unas ciencias sociales "críticas" y "comprometidas" con las cuestiones de cada tiempo. La investigación científica debe observar, respecto a la observación y análisis de los hechos, una objetividad o neutralidad valorativa, tan perfecta como humanamente sea posible. Sin embargo, en el objeto y orientación de sus investigaciones no debe ser ajena a los grandes problemas sociales que agitan en cada momento al mundo en que vivimos.

Desde un punto de vista personal, el requisito de objetividad no es fácil de cumplir en la práctica, porque el científico es un hombre y, como tal, tiene sus creencias y convicciones, de las que es imposible despojarse totalmente. Sin embargo, lo que está obligado es a evitar toda falta de objetividad consciente.

Se ha de tener en cuenta que la neutralidad valorativa no significa desprenderse de los propios valores, lo que es imposible, sino que exi-

ge únicamente atenerse a los hechos y que se los respete. Es decir, que en ningún momento se deben velar, adulterar o falsificar porque sean molestos o contrarios a las convicciones propias. Este modo de proceder es contraproducente. La realidad se venga. Las cosas no dejan de ser como son porque las disfracemos. En estos consiste fundamentalmente la ética de la investigación científica.

TIPOS DE INVESTIGACION SOCIAL

Los tipos de investigación social son distintos según se atienda a los siguientes aspectos de la investigación social:

a) su finalidad,
b) su alcance temporal,
c) su profundidad,
d) su amplitud,
e) sus fuentes,
f) su carácter,
g) su naturaleza,
h) su objeto,
i) el ambiente en que tienen lugar,
j) los tipos de estudios a que den lugar.

a) *Finalidad.*— Según su finalidad, se puede dividir la investigación social en, básica y aplicada. La primera tiene como finalidad el mejor conocimiento y comprensión de los fenómenos sociales. Se llama básica porque es el fundamento de toda otra investigación. Por el contrario, la investigación social aplicada, busca mejorar la sociedad y resolver sus problemas.

Consiste, de aquí su nombre, en la aplicación de los logros de la investigación básica, de la que por tanto depende, a los fines indicados.

De los cuatro oficios principales que cumple la investigación, conocer, explicar, prever o predecir y actuar, los dos primeros constituyen el objeto de la investigación básica, y los dos últimos son aplicaciones de ella y, por tanto, entran en el campo de la investigación aplicada. Ambas investigaciones están pues estrechamente vinculadas.

TIPOS DE INVESTIGACION SOCIAL SEGUN SU

- *fin*: básica / aplicada
- *alcance temporal*: seccional / longitudinal — retrospectiva / prospectiva / de panel / de tendencia
- *profundidad*: descriptiva / explicativa
- *amplitud*: microsociológica / macrosociológica
- *fuentes*: primarias / secundarias / mixtas
- *carácter*: cualitativas / cuantitativas
- *naturaleza*: documentales / empíricas / experimentales / doctrinales
- *marco*: de campo / de laboratorio
- *los estudios a que dan lugar*: piloto / evaluativas / informes sociales / de un caso / sondeos / encuestas / replicación
- *objeto sobre*: disciplinas / instituciones sociales / sectores sociales

En síntesis se puede decir que la finalidad de la investigación social en su conjunto, es el conocimiento de la estructura e infraestructura de los fenómenos sociales, que permita explicar su funcionamiento (investigación básica) con el propósito de poder llegar a su control, reforma y transformación (investigación aplicada).

b) *Alcance temporal.*– Conforme a este aspecto, la investigación puede referirse a un momento específico o a un tiempo único t, o puede extender su análisis a una sucesión de momentos temporales t_1, t_2. En el primer caso recibe el nombre de seccional. Se hace, por así decirlo, un corte perpendicular de una situación en un momento dado y se estudia su estructura.

En el segundo caso, se llama investigación longitudinal. También reciben el nombre de sincrónicas las seccionales y de diacrónicas las longitudinales. Aquí el corte es transversal, lo que permite estudiar la evolución del fenómeno en el período dado. Las investigaciones longitudinales se pueden subdividir en retrospectivas, en las que la serie de momentos estudiados se refieren al pasado, y prospectivas, cuando se refieren al presente y futuro.

Los estudios longitudinales se llaman también de panel, si en las investigaciones sucesivas se observan siempre los mismos individuos y de tendencia, si son diferentes.

c) *Profundidad.*– De acuerdo con este criterio, se pueden dividir las investigaciones sociales en descriptivas, explicativas y exploratorias. Las primeras son aquellas que conforme a la definición que da de ellas Hyman (1971, p. 102), tienen como objeto central "la medición precisa de una o más variables dependientes, en una población definida o en una muestra de una población".

Las investigaciones sociales explicativas son las que no solamente pretenden medir variables, sino estudiar las relaciones de influencia entre ellas, para conocer la estructura y los factores que intervienen en los fenómenos sociales y su dinámica.

d) *Amplitud.*– Por su amplitud, cabe distinguir las investigaciones microsociológicas y macrosociológicas. Las primeras son las que hacen referencia al estudio de variables y sus relaciones en grupos pequeños y medianos, mientras que las segundas se caracterizan por abordar dicho estudio respecto de grandes grupos o sociedades humanos.

e) *Fuentes.*– Según este criterio, se pueden distinguir las investigaciones sociales según se realice basándose en datos primarios o bien en datos secundarios. Las primeras son aquellas en que los datos o hechos sobre los que versan, son de primera mano, es decir, recogidos para la investigación, y por aquéllos que la efectúan.

Las segundas son, por el contrario, las que operan con datos y hechos recogidos por distintas personas y para otros fines e investigaciones diferentes. También existen investigaciones mixtas, que aplican a la vez datos primarios y secundarios.

f) *Carácter.*– Esta división hace referencia a dos enfoques históricos principales de la investigación social (Bruyn, 1972, págs. 26 y 22).

El primero, el cuantitativo, es el que centra de manera predominante la investigación social en los aspectos objetivos y susceptibles de cuantificación de los fenómenos sociales.

Este enfoque es defendido por el principio de Durkheim (1964, pág. 41), según el cual "la regla primera y más fundamental del método científico es considerar los hechos sociales como cosas".

El segundo, el cualitativo, es por el contrario, el que se orienta a descubrir el sentido y significado de las acciones sociales, según Max Weber, o lo que Znaniecki llamaba el coeficiente humanístico de dichos fenómenos.

Ambos tipos de investigaciones no se deben considerar como opuestos e incompatibles.

Aunque los dos tienen, sin duda, su campo más apropiado, el cuantitativo en los estudios de fenómenos y de grupos pequeños, y el segundo, en las encuestas referentes a grandes poblaciones, se debe procurar siempre que las investigaciones cualitativas sean también precisas y exactas en lo posible, y que las cuantitativas intenten penetrar y comprender los aspectos íntimos de los fenómenos que estudian.

g) *Naturaleza.*– Según su naturaleza, las investigaciones pueden ser

1) Empíricas, que trabajan con hechos de experiencia directa no manipulados.

2) Experimentales, que se apoyan en la observación de fenómenos provocados o manipulados en laboratorios o ambientes artificiales.

3) Documentales, que tienen como objeto directo la observación de fuentes documentales.

4) Encuestas, en las que los datos manejados proceden de las manifestaciones verbales o escritas de los sujetos observados.

A ellas se podría añadir, los trabajos doctrinales o filosóficos que tratan cuestiones puramente teóricas o de naturaleza supraempírica

h) *El objeto social al que se refieren.*– De acuerdo con este punto de vista, se pueden distinguir distintas clases de investigaciones sociales según la disciplina social a que se refieren, las instituciones sociales que estudian y los sectores sociales a que corresponden.

Por la disciplina, las investigaciones sociales pueden ser de Sociología, de Psicología social, de Antropología social y de Ecología social.

Por las instituciones que estudien se pueden referir a la familia, la economía, la empresa, el derecho, la política, el ocio y los *mass media*, la religión y la educación.

Por los sectores sociales a que correspondan cabe distinguir investigaciones sociales sobre el sector rural, el urbano, la estructura social, la cultura y sus manifestaciones, la desviación social y los problemas sociales.

i) *Según el marco en que tienen lugar.*– Se tienen las investigaciones sobre el terreno, o de campo, y las de laboratorio. Las primeras son las que se realizan observando el grupo o fenómeno estudiado en su ambiente natural y las segundas, en un ambiente artificial, cual es el laboratorio.

j) *Tipos de estudios a que dan lugar.*– Además de las clases de investigación indicadas, cabe mencionar las siguientes, que representan modalidades particulares de investigación social de interés por su frecuente aplicación: los estudios piloto; las investigaciones evaluativas; los informes sociales; los estudios de replicación; las encuestas y los sondeos y el método de casos.

Los estudios piloto se definen en el Diccionario de Ciencias Sociales del Instituto de Estudios Políticos, como investigaciones preliminares que preceden a las encuestas sociales, con el fin de perfeccionar y probar las técnicas empleadas.

Las investigaciones evaluativas tienen por objeto apreciar y enjuiciar el diseño, ejecución, efectos, utilidad y el grado en que alcanzan los objetos pretendidos los programas de acción social llevados a cabo en el campo de la sanidad, la educación, el ocio, etc., con el fin de corregir las deficiencias que pudieran existir e introducir los reajustes necesarios.

Los informes sociales se distinguen de las monografías por su mayor amplitud. En vez de restringirse a un solo aspecto social o sector, son estudios detallados de todos los aspectos de una comunidad, ciudad o comarca determinada.

El método de casos, *case studies*, es una especie de método biográfico en sentido amplio, que tiene por objeto recoger los datos que describen el proceso vital de una persona, una familia, una institución e incluso, una nación, utilizando frecuentemente documentos personales. Son estudios muy concretos, sobre un caso determinado, que se estudia en profundidad y en sus diversos aspectos.

Las encuestas son investigaciones que, como los informes, normalmente se extienden a amplios sectores de la población y pretenden obtener información sobre múltiples temas. Se diferencian de los informes sociales en que sólo utilizan técnicas de observación basadas en los distintos tipos de cuestionarios, y no toda clase de técnicas, como los informes.

Los sondeos de opinión se pueden considerar como un tipo especial de encuesta, caracterizado porque el ámbito de los temas sobre los que pretenden obtener una información, suelen ser mucho más reducidos y concretos que en las encuestas.

Por último, los estudios de replicación son los que se realizan para repetir investigaciones ya realizadas, en unos ambientes y condiciones nuevos. Tienen un gran interés científico.

Un descubrimiento científico sólo se puede considerar logrado, cuando ha sido confirmado por numerosas y repetidas investigaciones. (Ver ejercicio 1).

EJERCICIOS

1

Señalar razonadamente los tipos a los que pertenecen según: a) su finalidad; b) su alcance temporal; c) su profundidad; d) su amplitud; e) sus fuentes; f) su carácter; g) naturaleza y h) la clase o tipo especial de estudios a que dan lugar.

A) Las investigaciones patrocinadas por la Fundación Foessa, sobre la situación social en España.

RESPUESTA

Toda clasificación supone normalmente un cierto grado de ambigüedad. Por lo general, en la realidad no hay tipos puros. La clasificación, pues, de una investigación en un tipo, no se opone a que pueda presentar también ciertos rasgos de otro tipo.

a) Su finalidad es básica, en cuanto lo que se busca es conocer y analizar la situación social de España.

b) Su alcance temporal es seccional. Se pretendía conocer la situación social de España, en cada uno de los momentos que se ha realizado y no su evolución temporal.

c) Por su profundidad se pueden considerar predominantemente descriptivas. Expresan ante todo los rasgos y caracteres de la situación social española en sus diversos aspectos, aunque pueda tener algo de explicativa.

b) Su amplitud es macrosociológica. Se refieren a un grupo social grande, España.

e) Sus fuentes son mixtas, primarias y secundarias. Ha utilizado en parte encuestas hechas para los informes y también datos secundarios ya existentes, estadísticos y de otras investigaciones.

f) Por su carácter, es cuantitativa. Se refieren a los aspectos estadísticos cuantificables de la realidad social española, y no buscan una comprensión íntima o profunda de la misma.

g) Por su naturaleza mixta, a la vez empírica y documental.

h) Por su tipo es un informe social.

B) La investigación realizada por E. Durkheim sobre los factores sociales que influyen en el suicidio, con base en las estadísticas, de este fenómeno social en Europa.

RESPUESTA

Por su finalidad es básica. Durkheim pretende analizar y explicar el suicidio y no solucionar este problema social.

Por su alcance temporal se puede considerar longitudinal retrospectiva, en cuanto la base del estudio son las estadísticas del suicidio en Europa, no de un año, sino de muchos.

Por su profundidad es explicativa. Pretende explicar los factores que influyen en el suicidio. De hecho, es origen de importantes teorías al respecto, y en relación a la anomia y conducta desviada.

Por su amplitud es macrosociológica. Estudia un gran grupo social, gran parte de Europa.

Por sus fuentes, es claramente secundaria. No obtiene él los datos de su investigación, sino que utiliza las estadísticas existentes ya.

Por su carácter, es cuantitativa. Se refiere a los aspectos estadístico-cuantitativos del suicidio, y a su relación con factores sociales objetivos externos a los individuos.

Por su naturaleza, es documental.

Por su tipo se puede incluir en las monografías, por ser un trabajo sobre un tema bastante concreto, el suicidio que se estudia en profundidad desde un punto de vista sociológico.

C) La investigación efectuada por Max Weber en 1920, sobre la influencia de la ética protestante, en el origen y estudio del capitalismo liberal de los siglos XVIII y XIX, con base en datos históricos.

RESPUESTA

Por su finalidad es básica. Su objetivo no es sino estudiar la influencia de la ética protestante, en el origen del capitalismo moderno. Temporalmente, se puede cali-

ficar de longitudinal retrospectiva, porque estudia las manifestaciones y la evolución comparativa de dos fenómenos sociales históricos, la ética protestante y el espíritu capitalista. Por su amplitud es macrosociológica. Estudia el origen del capitalismo con referencia a toda Europa.

Por su profundidad es explicativa, porque, además de describir los fenómenos que estudia, lo que pretende es determinar si se puede considerar la ética protestante en alguna forma, motivo o causa de la otra, el espíritu capitalista.

Por sus fuentes es también claramente secundaria. Marx Weber se sirve de datos históricos, que ya existian.

Por su carácter, es cualitativa, porque estudia a un fenómeno único e intenta una comprensión íntima, cualitativa del mismo.

Por su naturaleza, es histórica-documental.

Por su tipo especial también se puede calificar de monografía, por ser el estudio en profundidad de un tema muy concreto.

D) Una investigación realizada por encargo de una casa de neumáticos, para indagar entre los camioneros la marca de neumáticos que utilizan, su satisfacción con ellos y los problemas que tenían al respecto.

RESPUESTA

Por su finalidad es aplicada. Tiene una finalidad práctica. Conocer la situación del mercado y los gustos de los consumidores de neumáticos, con fines productivos.

Por su alcance temporal es seccional. Se refiere a un momento dado, el de la investigación.

Por su profundidad, es sin duda principalmente descriptiva.

Por su amplitud es macrosociológica. Se refiere a todos los camioneros españoles, que sin duda son bastantes miles.

Por sus fuentes es primaria. Se basa en una encuesta hecha por los investigadores.

Por su carácter es cuantitativa.

Por tipo especial, o bien se puede considerar como encuesta o como sondeo.

E) Los estudios de Elton Mayo y sus colaboradores, por medio de experimentos sobre grupos de obreros de la Western Electric, para determinar las influencias en el rendimiento de los obreros y las relaciones sociales entre ellos y sus jefes.

RESPUESTA

Por su finalidad, se pueden calificar de aplicados, dada su orientación a una finalidad práctica: la solución de los problemas de rendimiento laboral.

Por su alcance, temporal, se puede llamar seccional.

Se refiere más que al estudio de la evolución del fenómeno, a la situación del mismo, en el momento en que se estudia.

Por su profundidad se puede calificar de explicativa, porque busca determinar los motivos y factores en que influyen en el rendimiento laboral.

De hecho, esta investigación tuvo una gran importancia práctica, porque fué el origen del movimiento en política social, llamado de las relaciones humanas, es decir de una política de la empresa orientada a dar al trabajo más ventajas y mejor tratamiento, con vista a mejorar su productividad.

Por sus fuentes, es primaria. Utilizaron datos obtenidos por ellos mismos, fundamentalmente mediante experimentos y entrevistas no dirigidas.

Por su amplitud se puede calificar de microsociológica. Los experimentos se hicieron con pequeños grupos de obreros, los que formaban distintos equipos o secciones de la empresa y a los que observaban mientras trabajaban, manipulando las condiciones de trabajo, por ejemplo, la iluminación.

Por su naturaleza, es experimental.

Por su carácter me inclinaría a considerarla cuantitativa. Más que al aspecto interno de los fenómenos sociales a ellos les interesaban los aspectos externos.

Por su tipo especial, es difícil de encuadrar en los tipos indicados. Más bien se debe considerar como experimento que es un tipo no incluido en la clasificación dada.

F) La investigación "Los hijos de Sánchez" realizada por O. Lewis con relación a una familia mejicana típica de la clase pobre, mediante varios años de convivencia con esta familia, observación de la misma y recolección de todos los detalles de la vida de sus miembros y de la familia, por entrevistas registradas en magnetófono.

RESPUESTA

Por su profundidad es descriptiva. Pretendía describir y describe la vida de los miembros de la familia. aunque pueda ser base de análisis explicativos.

Por su alcance temporal, es más bien longitudinal, en cuanto que describe la evolución vital de la familia.

Por su finalidad, es básica. Lewis no pretendía ninguna finalidad práctica inmediata con su estudio, sino estudiar científicamente la familia para conocer sus características.

Por su amplitud, es claramente microsociológica. Versa sobre un grupo pequeño: una familia.

Por sus fuentes, es primaria. Se basa en los datos observados y recogidos por el investigador.

Por su carácter, es cualitativa. No busca cuantificar, sino comprender el sentido y significado sociológico de las acciones sociales estudiadas.

Por su tipo se puede calificar como un estudio de caso. Es el estudio de un solo caso social muy típico.

Por su naturaleza, es empírica.

3. EL PROCESO DE LA INVESTIGACION SOCIAL

LOS ASPECTOS DEL PROCESO DE INVESTIGACION CIENTIFICA

Según se deriva de lo expuesto hasta aquí, la investigación científica es en sí, esencialmente, una actividad, pero no una actividad única y simple sino compleja o un proceso, constituida como tal por un conjunto de actuaciones sucesivas interrelacionadas.

Por eso, parece necesario exponer en qué consiste este proceso por lo menos en sus líneas esenciales, sobre todo si se tiene en cuenta que describe el contenido y las fases de la investigación científica y, por ello mismo, la pauta fundamental que señala las actuaciones básicas a desarrollar y el orden a seguir en dicha investigación.

No obstante lo anterior, hay que señalar que esta forma típica se ha de acomodar a los muy diversos objetos posibles de investigación científica, por lo que puede adoptar en ellos diversas modalidades dentro de este modelo típico.

En general se puede afirmar que la investigación científica es un proceso que tiende a conceptualizar la realidad objeto de investigación, es decir, a obtener conocimientos, ideas, representaciones intelectuales de la realidad, que sean expresión lo más exacta posible de ella y contribuyan a engrosar el acervo teórico de las ciencias.

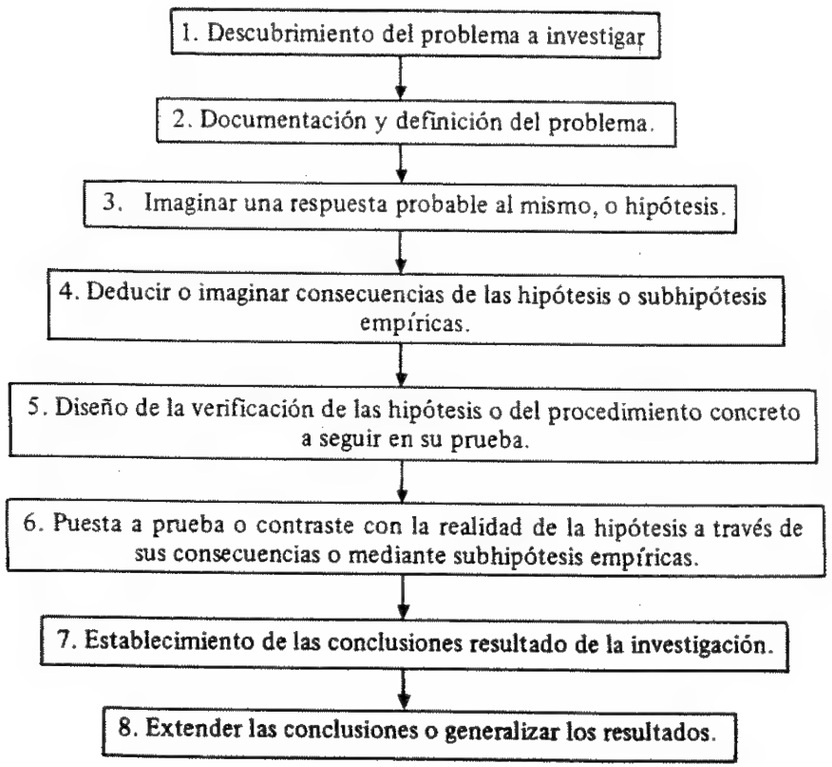
En este proceso global, se pueden distinguir tres aspectos a los que aquí se da el nombre de proceso metodológico, al primero, y lógico, al segundo. El primero enfoca la investigación desde el punto de vista de los pasos a seguir en la búsqueda de solución o respuesta al problema planteado. El segundo lo hace, desde el punto de vista de los elementos conceptuales que intervienen en la investigación científica y de su ilación lógica. El tercero, en fin, es el expositivo y consiste en la elaboración y redacción de los resultados de la investigación.

La exposición de los dos primeros procesos proporcionan, respecto a la investigación científica, la visión general de la marcha metodológica y lógica de la investigación científica en sus líneas fundamentales.

EL PROCESO METODOLOGICO

Se le puede aplicar este nombre porque la investigación científica, considerada desde este punto de vista, reproduce las etapas típicas del camino que hay que recorrer hasta llegar a la meta propuesta en la investigación. No es de extrañar que así sea, si se tiene en cuenta que, según la definición expuesta, la investigación implica la aplicación del método o camino científico a problemas concretos en el área de la realidad observable.

Las etapas del proceso metodológico de investigación científica que se pueden considerar típicas son, teniendo en cuenta a Bunge son las siguientes:



OPERACIONES DE LA INVESTIGACION SOCIAL

Las etapas del proceso metodológico comprenden diversas operaciones, que serán estudiadas en los capítulos de esta obra y se exponen ahora, en síntesis, para ofrecer una visión de conjunto de las mismas.

El problema a investigar es el origen concreto de la investigación y consiste en una pregunta o un interrogante sobre la realidad. Constituye también su objetivo o fin próximo, en cuanto que lo que se pretende lograr con la investigación es su solución.

El problema descubierto será en su inicio, normalmente, vago y abstracto y no estará convenientemente precisado, por lo que es necesario concretarlo, con toda claridad, en las operaciones subsiguientes.

Para realizar debidamente y con conocimiento de causa esta precisión, se requiere el estudio a fondo del tema de que se trate para informarse de las teorías o conocimientos científicos sobre él y de las investigaciones antes realizadas con relación al mismo. Esta información teórica se debe completar con otra empírica sobre la realidad o campo social a que se va a referir nuestra investigación, para obtener un conocimiento genérico de ella.

Efectuada esta labor informativa y de estudio, se debe proceder basándose en ella a enunciar el problema de forma concreta y con toda precisión y a fijar los objetivos de la investigación.

Respecto a la etapa siguiente la investigación científica reclama, por una parte, que en la búsqueda de solución al problema de investigación elegido, en lugar de proceder a la ventura, se trabaje ordenada e inteligentemente, imaginando las soluciones al problema más probables o hipótesis, a fin de proceder primero a su verificación. Estas hipótesis especifican el objeto de la verificación y la orientan en todas sus fases.

Por otra parte, el contraste de las hipótesis generales de la investigación, frecuentemente de carácter abstracto, se realiza normalmente no de modo directo sino mediante la imaginación y deducción de consecuencias empíricas muy concretas, a las que, si son inmediatamente verificables, se suele llamar subhipótesis.

Tratar con hipótesis y utilizar para su prueba consecuencias empíricas muy concretas, derivadas de ellas, que se puedan contrastar con la realidad fácilmente, es importante en toda investigación, aunque su carácter sea principalmente documental y cualitativo.

El diseño de la investigación debe especificar y planificar la forma concreta de realizar la prueba que aquella implica. No es otra cosa que la respuesta a los interrogantes de cómo se puede probar que la hipótesis es verdadera y qué pauta se ha de seguir en la recogida y tratamiento

de los datos. Debe estar en relación con las hipótesis y los objetivos de la investigación y decidir qué datos y sus tablas son necesarios.

Entre las operaciones que comprende se pueden destacar las siguientes:

- Especificar las variables y sus relaciones objeto de investigación y hacer aquéllas operativas, si son demasiado generales.
- Determinar qué otras variables, no objeto de la investigación, pueden influir en los resultados y prever los procedimientos a utilizar para su control.
- Precisar los datos sobre las variables investigadas necesarios y, en especial, determinar dónde obtenerlos, cómo recogerlos y el modo de tratarlos.

La realización de estas operaciones exige, respecto de dónde obtener los datos, delimitar en el espacio y en el tiempo el campo de observación y definir la población o conjunto de unidades de observación a que se va a extender la investigación. En el caso de que se opte por observar no todo el universo, sino una parte representativa o muestra de él, pide determinar, asimismo, el tamaño de la muestra necesaria y su error, formar su base y elegirla.

En cuanto a cómo obtener los datos, antes de empezar el trabajo de campo, reclama:

a) elección de las técnicas de observación a emplear de acuerdo con las variables empíricas que hayan de ser observadas y formular el diseño o plan de su empleo;

b) la construcción de los instrumentos para efectuar la recogida de datos si no están ya construidos;

c) su aplicación a una muestra muy reducida para probar su validez o pre-test.

Por último, el modo de tratamiento de los datos pide que se prevean las tablas de datos necesarias y los tipos de análisis estadísticos o de otro tipo adecuados.

Las operaciones básicas de la prueba diseñada son la observación, la clasificación y el análisis.

La observación en las ciencias sociales consiste en el examen o consulta de las unidades de observación para obtener los datos sobre las variables investigadas.

La clasificación es la agrupación sistemática, previa su codificación, de los datos recogidos referentes a cada variable y su presentación conjunta en tablas, bien separadamente o relacionados con los de otras variables.

El análisis busca hacer explícitos los rasgos, aspectos y propiedades de todo tipo que, con relación a las variables estudiadas, se derivan de las tablas en las que se condensa la clasificación. En la interpretación se determina la significación sociológica del análisis con inferencia de las conclusiones pertinentes respecto a la cuestión investigada.

Obtenidas las conclusiones se deben comparar con las hipótesis formuladas y la teoría fundamento de la cuestión investigada, integrando dichas conclusiones en ésta con las reformas y reajustes de la misma que sean precisos, y especificando las nuevas investigaciones a realizar.

Continuación del anterior es el proceso expositivo, que implica la elaboración de los materiales recogidos en la investigación y la exposición de sus resultados con el fin de comunicarlos a la comunidad científica y al público en general y posibilitar, de este modo, no sólo su conocimiento, sino la crítica y revisión de los mismos.

Hasta ahora no ha sido reconocido explícitamente este proceso expositivo como un elemento esencial de la investigación científica, salvo casos aislados. Sin embargo, el proceso de exposición es algo esencial y necesario, en cuanto da forma a los materiales recogidos y los resultados obtenidos en la investigación, que, en otro caso, no pasarían de constituir una masa informe de datos, tablas e interpretaciones.

Este proceso se concreta en el informe donde se hará una exposición del método, proceso de la investigación y de sus resultados, con inclusión de las tablas formadas. (Ver ejercicio 2A). Se estudia específicamente en mi obra *"Tesis doctorales y trabajos de investigación científica"* (Paraninfo, 1996, 4ª edición).

EL PROCESO LOGICO

Este proceso lógico es paralelo al metodológico pero en él se tienen en cuenta, no las etapas seguidas en el camino de la investigación, sino los elementos conceptuales que intervienen en ella.

En el proceso lógico se pueden distinguir dos subprocesos de movimiento inverso: de verificación y de teorización. En el primero, en síntesis, la investigación parte de la teoría existente y desciende a la realidad empírica. Por el contrario, en el segundo la investigación parte de la realidad y asciende a la teoría sintética.

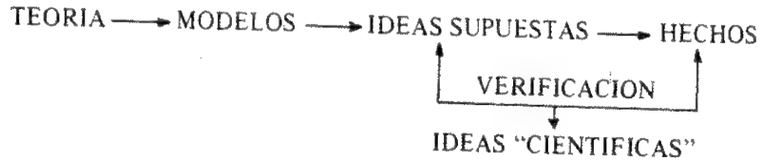
Adviértase cómo según este esquema, el proceso de investigación empieza con ideas y termina en ideas. A la realidad, a los hechos, hay que ir con ideas, los hechos de por sí, no dicen nada. Igualmente, no se pueden quedar uno en los hechos, éstos solos y aislados, no tienen significación científica.

EL PROCESO DE VERIFICACION

El proceso de verificación, como su nombre indica, es un proceso probatorio, de prueba en la realidad de la teoría.

Sus elementos fundamentales: la teoría, los modelos, la hipótesis, la realidad, las ideas "científicas" y la verificación que relaciona las ideas y los hechos.

Su esquema podría ser el siguiente:



TEORIA

La palabra teoría se deriva etimológicamente de los términos griegos *theoria* que significa visión, vista, contemplación, especulación mental y *theoreo* con la significación de mirar, observar, contemplar.

De acuerdo con este origen etimológico, con el término teoría, tomado en su sentido más general, se designa en los idiomas modernos toda concepción racional que intenta dar una explicación sobre cualquier realidad.

Desde un punto de vista científico las teorías se pueden definir de acuerdo con el Diccionario de Ciencias Sociales (1976), como un conjunto de proposiciones conectadas lógica y ordenadamente que intenta explicar una zona de la realidad mediante la formulación de las leyes que la rigen.

En las proposiciones que forman las teorías, se pueden distinguir varios grupos.

El primero es el constituido por las ideas "científicas", resultado de la prueba con los hechos de las hipótesis o ideas supuestamente científicas.

El segundo grupo es el de las ideas deducidas lógicamente de enunciados contrastados empíricamente.

El tercero, está constituido por principios que tienen un origen axiomático, es decir, constituyen verdades básicas en una ciencia que se aceptan sin demostración.

Por último, en el cuarto se pueden incluir los principios racionales no axiomáticos y no verificados aún empíricamente.

La consecuencia de esta distinción es importante. Hay que tener en cuenta que no todo el contenido de las teorías y por tanto de la ciencia, tienen un fundamento empírico inmediato y directo. Gran parte de este contenido es lógico o deductivo. Es así, porque las teorías intentan explicar lógica o racionalmente los resultados de la observación empírica.

Las teorías son el resultado final de la investigación científica y las que constituyen en último término la ciencia.

Todas las teorías existentes, referentes al campo que abarca una ciencia, forman objetivamente esta ciencia. La ciencia, pues, está formada esencialmente por teorías. No es por tanto otra cosa que el conjunto de teorías referentes al ámbito que abarca. Todas las teorías de una ciencia constituyen el cuerpo de conocimientos que se suponen válidos, existentes en una época determinada sobre el sector de la realidad propio de la ciencia de que se trate.

Lógica y lingüísticamente, las teorías no son otra cosa que conjuntos de enunciados conectados lógica y ordenadamente entre sí, es decir, conjuntos de expresiones (afirmaciones o negaciones) referentes al sector de la realidad objeto de la ciencia.

La teoría es un elemento muy importante de la investigación social. En cierto modo, se puede decir que es su origen, su marco y su fin. Su origen porque son fuente de nuevos problemas e hipótesis. Su marco, porque proporcionan el sistema conceptual que se aplica a la observación, clasificación y sistematización de los datos de la realidad. Son su fin, porque la investigación debe desembocar en teorías cada vez más perfectas.

Todo ello explica que como destaca M. Bunge (1972, p. 413) en su obra "La investigación científica" sea "una peculiaridad de la ciencia contemporánea el que la actividad científica más importante, la más profunda y la más fecunda se centre en torno a teorías y no en torno a la recolección de datos, las clasificaciones de los mismos o hipótesis sueltas".

Como señalan Mannheim y Rich (1988, 26) las teorías deben reunir los rasgos de ser contrastables, lógicamente correctas, comunicables o comprensibles, generales y parsimoniosas, es decir no enrevesadas y profusas.

Sin embargo, T.S. Kuhn (1975, p. 128) disiente de esta posición de Popper y estima que "una teoría científica se declara inválida sólo cuando se dispone de un candidato alternativo que ocupe su lugar".

La verificación, entendida de la manera indicada, como prueba de no oposición a la realidad, puede dar lugar, alternativamente, a dos resultados: la confirmación en un caso concreto o su refutación, y se puede distinguir según se refiera simplemente a hipótesis o a teorías.

La verificación o prueba de hipótesis se realiza, directamente, a través de todo el proceso de investigación. Si el resultado de la investigación es positivo quiere decir que la hipótesis ha quedado confirmada por los hechos de la realidad investigada. Por el contrario, si es negativo, hay que rechazar la hipótesis por lo menos tal como ha sido formulada y, reformarla, en el caso de que comprobada anteriormente, se ajuste a otras realidades pero no a ésta.

En cuanto a la verificación o prueba empírica de teorías, tiene lugar principalmente de forma indirecta. (Cfr. Popper, 1973, 32 y ss) mediante la derivación de la teoría de conclusiones o hipótesis que impliquen predicciones específicas de lo que debe ocurrir o no en la realidad. La confirmación o refutación de una hipótesis, en la investigación o experimento que se realice, supondrá un mayor apoyo empírico de la teoría o la necesidad de su modificación, respectivamente.

EL PROCESO DE TEORIZACION

Este proceso empieza donde termina el de verificación, en los hechos, en la realidad. Sigue un movimiento opuesto a éste. El proceso de verificación es descendente, y por así decirlo, deductivo. El proceso de teorización es ascendente, inductivo. Se eleva de los hechos gradualmente a la teoría.

De los hechos se obtienen datos, con los datos se forman conceptos y enunciados, con los enunciados se articulan teorías que pasan a engrosar el acervo de la ciencia.

Tres son pues los elementos de este proceso: la obtención de datos, la formación de proposiciones y la articulación de teorías. Como ya he hablado de las teorías, me referiré a los restantes.

Los datos son los materiales más simples de la investigación científica. No son los hechos, sino expresiones sobre estos hechos o afirmaciones y negaciones que expresan aspectos o características concretas de los hechos.

En su origen, los datos son simples y concretos; se refieren siempre a una dimensión de una variable en una unidad de observación determinada. Datos por ejemplo son: El encuestado X es hombre y el Y tiene 25 años.

Los datos son esencialmente descriptivos. Establecen simplemente la ausencia o presencia de la dimensión a que se refieren. Excluyen, por tanto, toda inferencia o deducción, toda comparación y el establecimiento de relaciones. Se puede decir, por ello, que son el resultado mecánico de la observación.

Agrupando, relacionando y analizando los datos, se obtienen generalizaciones empíricas. Por ejemplo un 85% de los encuestados varones es favorable al ingreso de España en el Mercado Común. Luego, además de formarse conceptos científicos, si ello es necesario, se ve si las hipótesis son consistentes con la realidad investigada. En caso afirmativo, las hipótesis se transforman en enunciados científicos.

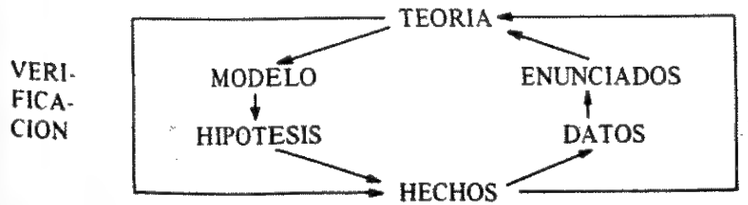
Las generalizaciones empíricas, si aparecen confirmadas por muchas investigaciones y representan regularidades válidas, por lo general, en toda población, reciben el nombre de leyes científicas, en cuanto establecen uniformidades que afecta a todo conjunto o universo.

Las leyes científicas en las ciencias sociales presentan generalmente carácter estadístico. Establecen únicamente que un porcentaje determinado de la población reune la característica o relación a que se refiere la hipótesis investigada.

En consecuencia, no proporcionan la certeza de que dado un individuo cualquiera del conjunto, éste poseerá la característica o relación investigada, sino únicamente una probabilidad mayor o menor de que así sea.

Los enunciados científicos, tanto individuales como generales o leyes, interrelacionados entre sí, forman las teorías. Pero las leyes se diferencian de las teorías en que enuncian hechos o fenómenos y no los explican como aquéllas. Se distinguen a su vez de los hechos singulares por su generalidad. Por ello, en las leyes universales desaparece toda referencia a un objeto, lugar y tiempo concretos.

Según todo lo expuesto, el esquema completo del proceso de investigación es el siguiente:



CARACTER DEL PROCESO DE INVESTIGACION SOCIAL

De lo expuesto más arriba se deriva claramente el carácter conceptual del proceso de la investigación social no solo porque su origen es de naturaleza conceptual, las teorías previas, sino porque su fin es la conceptualización de la realidad y debe desembocar en nuevas teorías. Por otra parte, todos los elementos indicados distinguidos tanto en el proceso de verificación, excepto los hechos, como en el de teorización: teorías, modelos, hipótesis, datos enunciados, y leyes tienen carácter conceptual pues son enunciados o están formados por éstos.

Y no puede ser de otra forma, ya que la investigación social se identifica con el proceso de conocimiento científico social y por tanto es una forma de conocimiento y una actividad intelectual o conceptual.

Este carácter conceptual implica también el carácter lingüístico del proceso de investigación social en cuanto que los conceptos precisan para su expresión, como veremos, de una forma lingüística. En consecuencia, es obvia la importancia de la lógica y de la lingüística en la investigación social, en la que el lenguaje constituye, además, un elemento fundamental de la vida social.

EL PROCESO DE INVESTIGACION Y LOS CONTEXTOS DE DESCUBRIMIENTO Y JUSTIFICACION

El proceso de investigación social y también el de investigación científica en general, de ninguna manera se reducen a reglas ni siquiera a la aplicación de técnicas y fórmulas lógicas o matemáticas.

Al contrario, exige igualmente y de manera necesaria una gran dosis de intuición y de imaginación, actividades mentales que escapan a la lógica y a la matemática y respecto de las cuales no hay métodos ni técnicas que permitan producirlas a voluntad en cualquier momento que se precisen, así como tampoco una pedagogía que pueda enseñar a obtenerlas aunque si pueda estimularlas hasta cierto punto.

Estos dos aspectos del proceso de investigación, por una parte, la utilización de técnicas y fórmulas lógicas y matemáticas, y por otra, la intuición y la imaginación, fueron puestos de relieve por Hans Reichenbach (1973, 240) con su distinción de los contextos de descubrimiento y de justificación.

Respecto a las operaciones del proceso de investigación, se puede afirmar que al primer contexto, el de descubrimiento, pertenecen principalmente, las que se pueden llamar operaciones básicas a preliminares de la investigación, descubrimiento del problema a investigar, formulación de hipótesis, operacionalización de variables, etc.

En cambio, al segundo, el de validación o justificación, corresponden todas las operaciones del proceso de verificación propiamente dicho: observación o recogida de datos, clasificación y análisis.

Es obvio, que esta distinción no se puede tomar de modo tajante, pues ni las operaciones básicas preliminares indicadas excluyen de manera absoluta toda actividad lógica o reglada ni, al contrario, se puede decir que la observación, la clasificación y el análisis, no se necesitan para nada la intuición y la imaginación.

Desde el punto de vista de la actividad de la razón, en el contexto de descubrimiento se trata de un proceso intelectual principalmente de invención y concreción. Por el contrario en el contexto inductivo de justificación predomina un proceso intelectual de abstracción, preferentemente lógico, que infiere consecuencias y establece síntesis.

EJERCICIOS

2 (A)

Supuesto que se nos ha encargado realizar una investigación sociológica sobre las opiniones de la población de una comunidad autónoma española respecto a la objeción de conciencia ante la obligación del servicio militar, indicar el tipo de investigación que se crea más adecuado para realizarla y las operaciones básicas que habría que efectuar para llevarla a cabo.

RESPUESTA

El tipo de investigación más adecuado sería la encuesta, en cuanto constituye el procedimiento básico utilizado en la investigación social para conocer las opiniones de grandes poblaciones.

Las operaciones necesarias para llevar a cabo la investigación serían las siguientes:

Selección del problema y su definición. El problema está fijado en el enunciado. Su definición exige realizar una labor previa de documentación sobre las teorías existentes y las investigaciones ya realizadas referentes al tema propuesto, que permita concretar los objetivos del estudio y los aspectos de la cuestión investigada.

Formulación de hipótesis. Al tratarse de una investigación de tipo descriptivo su enunciado no implica hipótesis explícitas generales, salvo la implícita de que por algún motivo no señalado se trata de una cuestión que interesa conocer en dicha comunidad.

No obstante, se podrían formular hipótesis tales como que la opinión favorable a la objeción de conciencia será mayor en los jóvenes que en la gente mayor; en las personas con nivel de estudios alto que en las de bajo; en las ciudades que en el campo.

Elección de la técnica de investigación. Al tratarse de una encuesta se puede realizar por cuestionario escrito y por entrevista, dependiendo la decisión sobre este punto del presupuesto, objetivos y medios de que se disponga.

Especificar las VV. En el enunciado sólo se señala una variable, opiniones sobre la objeción de conciencia ante la obligación del servicio militar, de carácter general, que es necesario especificar y operativizar mediante una serie de preguntas concretas que versarían, por ejemplo, sobre el conocimiento del tema por los encuestados; sus opiniones sobre la paz, el desarme, el ejército, el servicio militar, la admisión y la forma de plantear la objeción de conciencia, las consecuencias de la misma, servicios sociales sustitutorios, tiempo, caracter, lugar y, en fin sobre los motivos de las opiniones.

Dónde obtener los datos. Respecto a este punto habría que delimitar geográficamente la comunidad autónoma, especificar el tiempo de la encuesta, y definir qué personas se considera forman parte de su población.

Como la población sería en todo caso muy amplia, sería preciso trabajar con una muestra, en la que estuvieran representados todos los estratos principales de la población, según sexo, edad, estudios, clase social, preferencias políticas.

Obtención de los datos. Exige la construcción del cuestionario, su prueba o pretest y su aplicación a la muestra seleccionada.

Tratamiento de los datos. Reclama la codificación de los datos obtenidos, su agrupación en tablas que relacionen las opiniones sobre el tema, por ejemplo, con el sexo, la edad, la educación, las simpatías políticas, etc., y la realización de los tipos de análisis pertinentes. Por último, quedaría la interpretación de los resultados obtenidos, el establecimiento de conclusiones y su generalización.

Sección

OPERACIONE
BASICAS PRELIMINARI

4 . DETERMINACION DEL PROBLEMA DE INVESTIGACION

La investigación en concreto comienza con la determinación del problema a investigar.

La cuestión a investigar puede ser encargada por un organismo, entidad o persona, o puede ser elegida por el propio investigador.

En este último caso, la operación de la determinación del problema a investigar, consiste en la actividad, mediante la cual el sociólogo especifica de un modo concreto el tema sobre el que va a versar el trabajo científico que piensa emprender. Se trata de una actividad compleja que supone y exige la respuesta a dos interrogantes: qué investigar y buscando qué. Requiere establecer no sólo el área o fenómeno social que se va a estudiar (qué investigar), sino también precisar qué es lo que se intenta saber o descubrir de dicho fenómeno (buscando qué).

Dado que tanto los fenómenos potencialmente investigables como las facetas y cuestiones que plantean son múltiples, la determinación del problema a investigar comporta dos elecciones: la primera del campo de estudio y la segunda de sus vertientes o de las cuestiones que plantee. En consecuencia, no basta para emprender una investigación social y definir el problema de investigación decir voy, por ejemplo, a estudiar el deporte, sino que es preciso concretar los aspectos que se intenta conocer, si se trata de un estudio descriptivo, o las cuestiones a las que se pretende obtener una respuesta, si es un estudio explicativo que busque descubrir los motivos desconocidos de la dificultad planteada.

El punto de partida de la investigación sociológica, como en general de toda investigación científica, es la determinación del problema a investigar. Es evidente que al emprender una investigación, la operación primaria y básica es decidir qué se va a investigar y bajo qué aspectos. Si falta esta determinación, pueden darse experiencias aisladas e inconexas acaso importantes, pero no puede hablarse de verdadera investigación científica en sentido estricto. Esta exige, como elemento esencial, proceder de forma sistemática y metódica en el estudio de un problema específico, previamente elegido.

CONDICIONES DEL PROBLEMA DE INVESTIGACION

Respecto a este punto, parece clara la necesidad práctica en que se halla el sociólogo de conocer las condiciones que debe reunir el problema a investigar para ser factible y tener valor científico. Este conocimiento le evitará el peligro de empeñarse en estudios inviables o sin interés científico alguno.

Entre estas condiciones se pueden destacar las siguientes, que pueden servir para contrastar si un problema elegido es o no adecuado:

1. Si la investigación ha de ser sociológica es preciso, en primer lugar, que el problema tenga tal carácter.

2. El problema a investigar no ha de ser vago y genérico, sino que debe ser concreto y estar formulado lo más precisamente posible.

"Muchos proyectos de investigación –escribe Lundberg (1949, página 58)– fracasan principalmente por falta de objetivos claramente definidos. Por ejemplo, después de un año de investigación sobre muchos temas, tales como 'vida de un niño', 'mujeres en la industria', 'el periódico', resultó que si bien llegó a reunirse gran cantidad de 'material interesante', prácticamente no pudo ser respondida ninguna cuestión de interés general."

3. La sociología pertenece al grupo de ciencias positivas referentes a la realidad, a lo que es, y no a lo ideal o a lo que debe ser. Por ello un problema a investigar que plantee juicios de valor sobre lo que es mejor o peor no es adecuado para una investigación sociológica empírica, aunque verse sobre materias sociales.

4. La cuarta condición es una consecuencia de la anterior. El carácter real, empírico, de la sociedad exige que los problemas que se planteen puedan ser objeto de observación y experimentación, o de algún modo de contraste y verificación con la realidad. De no ocurrir así, pertenecerá tal problema al campo de la filosofía social o de la moral, pero no propiamente al de la sociología.

5. Las ciencias positivas buscan el establecer generalizaciones y regularidades basadas en la observación de los hechos. En consecuencia, los problemas de investigación que se planteen no deben referirse a casos únicos, aislados, sino que deben ser representativos y susceptibles de generalización en el área a que se refieren.

6. En fin, el problema elegido debe representar alguna novedad. Según esto, no son admisibles investigaciones sobre fenómenos ya conocidos y estudiados o cuestiones ya resueltas, si no suponen algún enfoque o punto de vista nuevo que pueda significar un avance o desarrollo respecto a las metas ya conseguidas. (Ver ejercicio 2B).

ORIGEN DEL PROBLEMA DE INVESTIGACION

Con relación a este punto, se puede distinguir el origen concreto del problema de investigación y su origen genérico.

El origen concreto coincide con la cuestión del hallazgo del problema de investigación, o proceso a seguir en su determinación, que se examina a continuación.

Genéricamente se deben diferenciar el origen psicológico y el lógico.

Desde un punto de vista psicológico, los problemas de investigación se derivan de la curiosidad o de la necesidad. Lo que inclina a la persona humana a plantearse problemas y enfrentarse con su solución, es tanto la curiosidad: el afán de encontrar una explicación a lo desconocido y de conocer más y mejor el mundo que le rodea, como la necesidad, es decir, la urgencia de encontrar remedio a sus limitaciones y a las dificultades que encuentra en el medio en que vive.

Desde un punto de vista epistemológico, el origen del problema se halla en la limitación de nuestros conocimientos o ignorancia y sobre todo en el error.

Todo lo que el hombre desconoce o ignora es potencialmente un problema que puede ser objeto de investigación por él, con el fin de llegar a su comprensión. Como la ignorancia del hombre es ilimitada, también son los problemas de este tipo. Es más, el conocimiento los amplía en cierto modo en cuanto deja al descubierto nuevos interrogantes antes no advertidos.

En cuanto al error, es una de las fuentes más típicas de problemas de investigación. Frecuentemente, las ideas que creemos son ciertas, chocan con los hechos o son desmentidas por la realidad, lo que pone de manifiesto que son erróneas por lo menos en alguna medida. Es obvio que entonces se plantea el problema de resolver esta discordancia entre las ideas y la realidad, de hallar una explicación racional a la misma que puede y debe ser el origen de una investigación. En este sentido, se ha de prestar especial atención como fuente de nuestras investigaciones a todo hecho anómalo, imprevisto y estratégico, que se sale de lo normal o común y al que no se puede encontrar explicación en las teorías conocidas, el cual recibe el nombre de *serendipity* (v. Merton, 1970, 113 y ss.).

PROCESO DE DETERMINACION DEL PROBLEMA DE INVESTIGACION

El proceso a seguir en la determinación del problema de investigación es una cuestión, sobre todo en su aspecto básico de hallazgo y discernimiento de problemas de interés para la investigación, de visión y agudeza intuitiva del investigador, en la que no existen reglas fijas y sirven de poco las fórmulas previas.

A lo más que llegan los autores en este punto es a proponer algunos consejos y a formular algunas reglas genéricas al respecto. Por ejemplo, en el libro de Asti Vera (1968, pág. 101) se hacen las siguientes recomendaciones para encontrar problemas de investigación significativos:

– Atender a las incongruencias y contradicciones en los asuntos polémicos y a las conclusiones no demostradas.

– Seguir las huellas y sugestiones de lecturas, conferencias o de la simple reflexión.

– Convertirse en un erudito en una o más especialidades.

– Ejercitar el ingenio en todo obstáculo que se presente.

– Leer, escuchar y trabajar con sentido crítico.

– Iniciar una investigación y estar atento a los problemas que de ella surjan.

– Persistir en el examen de problemas ya delimitados.

También se debe tener en cuenta que la elección del problema de investigación debe ser un descubrimiento, como señala Gian A. Gilli (1971, p. 45 y ss.), y no una simple elección especulativa.

Debe ser un descubrimiento en el sentido de:

1. Que debe surgir del contacto íntimo con la realidad social y del conocimiento profundo de la teoría científica.

2. Que debe perfilarse y completarse a lo largo de todo el proceso de investigación.

Determinado pues, inicialmente, el problema a investigar, resultará que, tendremos, por lo general, especificado únicamente el qué investigar.

Hace falta, por tanto, precisar también inicialmente, es decir, sin perjuicio de los ajustes a que pueda dar lugar la investigación, el buscando qué, o sea, los objetivos concretos de la investigación.

Para efectuarlo con la mayor garantía científica, será de gran utilidad realizar en este momento las operaciones ya indicadas de la investigación:

a) El estudio de la teoría referente al tema elegido y de las investigaciones empíricas similares.

b) La observación global del campo social a investigar.

Para realizar la primera de estas operaciones es aconsejable examinar primero las enciclopedias especializadas en ciencias sociales y los tratados sobre la materia en cuestión, así como consultar a especialistas.

Después se revisará la bibliografía, especialmente la más destacada sobre el asunto, tomando en fichas notas de todos los puntos que nos interesan.

En cuanto a la segunda operación, haré referencia a ella al tratar de la observación directa simple.

Es indudable que la realización de estas operaciones después de hallado el tema a investigar, nos descubrirá el alcance y las implicaciones teóricas y prácticas que él mismo presente, lo cual indudablemente nos será de gran utilidad para precisar los fines de la investigación, lo que debemos buscar en ella.

Estos fines deben ser conocidos por todos los componentes del equipo investigador, pues deben ser tenidos en cuenta en todas las operaciones de la investigación.

TIPOS DE PROBLEMAS DE INVESTIGACION

Si bien es verdad que no se pueden dar reglas precisas para la búsqueda y hallazgo de temas de investigación sociológica significativos, sin embargo sí se pueden indicar los tipos más comunes en los que se encuadran, por lo general, tales trabajos. Esta enumeración es, sin duda, útil como orientación sobre las modalidades principales que suelen presentar los estudios que se emprenden en este campo y como sugerencia, en consecuencia, de las opciones más importantes a nuestra disposición cuando se trata de abordar un estudio empírico.

Los tipos a los que se pueden reducir los trabajos empíricos de investigación sociológica, además de los que representan los tipos de investigaciones expuestos en el capítulo segundo, son básicamente los siguientes:

En primer lugar se encuentran los descriptivos. Todo grupo, comunidad, institución cultural, fenómeno, acontecimiento, e incluso, objeto de interés social puede ser materia de un estudio descriptivo, en el que se analicen todos o algunos de sus principales elementos y caracteres de interés sociológico.

En segundo lugar se hallan los estudios de tipo comparativo. Simultáneamente o con referencia al mismo período de tiempo se pueden comparar entre sí todas las comunidades, grupos, culturas, fenómenos, acontecimientos y objetos distintos que pueden ser objeto de estudio descriptivo. La importancia de la comparación es tal que se puede afirmar que, salvo los estudios puramente descriptivos, toda investigación científica supone alguna forma una comparación.

En tercer lugar, figuran los trabajos sobre relaciones entre características, variables en términos técnicos, de los fenómenos y grupos sociales. En este sentido, es posible relacionar, lo que constituya una forma de comparación, el grado y signo de aparición y variación conjuntas, de las distintas propiedades que caracterizan a los fenómenos sociales. Por ejemplo, se puede investigar la relación entre el aislamiento social y las enfermedades mentales, el grado de religiosidad y la frecuencia de ir al cine, la asociación entre el sexo y la práctica del deporte, el nivel educativo y las actitudes políticas, la edad y el tipo de lecturas preferido. La enumeración de casos posibles sería, como en los casos anteriores, prácticamente interminable.

En cuarto lugar cabe destacar los estudios de opiniones y actitudes en los grupos sociales más diversos. Estos estudios pueden versar sobre cualquier tema, acontecimiento o persona de interés público.

En quinto lugar hay que mencionar los estudios que buscan bien determinar las causas o los motivos que explican los fenómenos sociales o bien precisar los efectos sociales de situaciones nuevas, de transformaciones sufridas, de innovaciones introducidas o de cualquier otro factor que haya actuado en un grupo, fenómeno o situación sociales dados. Por ejemplo, se pueden estudiar los efectos de una acción educativa, de una campaña publicitaria, de una inundación, de la puesta en práctica de un programa de desarrollo, o bien el porqué de la delincuencia juvenil y de la drogadición. En todos estos casos se trataría de estudios explicativos en cuanto tratan de determinar los efectos de ciertas causas, o al revés, las causas de ciertos efectos.

Otro tipo importante de investigación posibles está formado por las diversas modalidades de investigaciones empíricas, que versan sobre teorías científicas anteriores. A este respecto cabe, y es muy frecuente, la verificación de teorías admitidas con referencia a grupos, situaciones, tiempos o lugares nuevos; la comprobación de la posibilidad de extensión de teorías nacidas dentro de una ciencia social determinada a otra: por ejemplo, de la psicología social a la educación y viceversa; de la política a la economía, y al revés, etc. Incluso se puede dar, y de hecho aquí se encuentra el origen de importantes escuelas sociológicas, el tras-

plante de teorías del campo de las ciencias físicas y naturales al campo social. Por otra parte, se dan a veces conflictos entre distintas teorías existentes, o bien los hechos pueden demostrar la inadecuación total o parcial con la realidad de ciertas teorías admitidas o la existencia de lagunas en ellas, defectos y contradicciones éstos que se debe procurar esclarecer mediante nuevos estudios empíricos. Además, existen teorías que tienen un origen meramente deductivo, lo que reclama necesariamente su contraste con la realidad para ver su adecuación a ella.

El último tipo importante de estudios a señalar es el constituido por los dirigidos a solucionar problemas prácticos concretos que se plantean en la realidad social. Consisten en la aplicación de las conquistas y logros de la sociología científica a la solución de las cuestiones concretas de los más diversos sectores sociales, y especialmente relativas a los problemas sociales clásicos, tales como la vivienda, el urbanismo, el alcoholismo, las drogas, la prostitución, la contaminación, la congestión del tráfico, los subnormales, etc. (Ver ejercicios 5,6).

ASPECTOS DE LOS PROBLEMAS DE INVESTIGACION

La determinación de los objetivos de la investigación se puede enfocar desde otro punto de vista. Este es la especificación de los aspectos que ofrece el tema a investigar.

Por lo general, una reflexión atenta y reposada sobre el origen, naturaleza, carácter, funciones, elementos, historia, evolución y relaciones de la cuestión descubierta revelará la existencia de diversos aspectos en el problema elegido, susceptibles de ser tratados en conjunto o separadamente.

Al realizar este estudio de los aspectos del tema a investigar, se habrá ido perfilando la idea inicial del mismo, así como aclarando los objetivos perseguidos en el trabajo. También se obtendrá por lo general la convicción de la imposibilidad práctica de abarcar todos los aspectos que hemos descubierto y la necesidad de restringirse a alguno de ellos solamente.

Nunca se insistirá bastante en la necesidad de reducir prudentemente los objetivos de la investigación.

De aquí que sea preciso elegir, entre todas las facetas que presente el problema, según el análisis de éste realizado, solamente las que se crean más interesantes, significativas y adecuadas a la preparación, inclinación y medios del investigador. Con ello el objetivo de la investigación y el problema estudiado quedarán definidos concretamente.

Tampoco existen normas específicas a seguir en este paso o fase del proceso de referencia. Sin embargo, se pueden tener en cuenta como orientación en esta tarea los siguientes aspectos fundamentales, desde los que, por lo general, se pueden enfocar los fenómenos sociales:

a) Descriptivo: ¿Cuáles son sus elementos, órganos, aspectos y factores con una significación social?

b) Estructural: ¿Cómo están interrelacionados estos elementos y órganos entre sí?

c) Funcional: ¿Qué función cumplen los distintos órganos en el conjunto y respecto a la sociedad general, cómo están relacionadas unas funciones con otras y cuáles son sus aspectos funcionales y disfuncionales?

d) Demográfico: ¿Cuáles son los rasgos demográficos del elemento humano que forma en su caso el objetivo de investigación y qué influencia social tienen?

e) Ecológico: ¿Cuáles son las condiciones ambientales físicas en que se desarrolla el problema a investigar y cuál es su influencia social?

f) Conflictivo: Análisis de los conflictos producidos con relación al fenómeno en cuestión y estudio de sus motivos y efectos sociales.

g) Evolutivo: ¿Qué transformaciones sociales ha sufrido en el tiempo y su relación con el cambio social general?

h) Histórico: ¿Cuáles son los caracteres sociales peculiares que ha presentado a lo largo del tiempo?

i) Crítico-dialéctico: ¿Cuáles son las fuerzas e intereses sociales y económicos que actúan respecto al fenómeno de que se trate, y cuáles son los conflictos y tensiones que producen?

j) Genético: ¿Cuál es su origen social?

k) Cultural: ¿Cuáles son las normas, valores, creencias, realizaciones, técnicas y artísticas que lo caracterizan?

l) Proyectivo: ¿Cuál es su probable evolución futura?

Del mismo modo, también puede ser útil no perder de vista las relaciones más comunes y generales de los fenómenos sociales, tales como a) con las condiciones físicas y biológicas de lo social: clima, geografía, región, edad, sexo, raza, estado, salud; b) con sus condiciones culturales: normas, actitudes, valores, opiniones, creencias; c) con las condiciones personales de los miembros de la sociedad: status, rol, clase social, ocupación, situación económica, nivel cultural; d) con las instituciones y grupos sociales fundamentales: familiares, educativas, políticas, reli-

giosas, económicas y recreativas; e) en fin, con las condiciones dinámicas de lo social: cambio, conflicto y problemas sociales.

Con base en esta enumeración u otra similar se pueden enunciar, dado un tema a investigar, los aspectos generales, que, en principio, se pueden estudiar en él.

El hecho de tratarse de una faceta de la investigación científica no sujeta a reglas fijas, y que, en realidad, sólo se puede resolver en cada caso con referencia a las condiciones concretas de una situación de hecho, hace que no sea factible proponer ejercicios sobre la determinación de problemas a investigar que lo sean verdaderamente.

Por ello los ejercicios que siguen versan únicamente, como aplicación de las ideas expuestas en la introducción, sobre aplicación de las condiciones que deben reunir los posibles temas objeto de investigación, la determinación de aspectos susceptibles de estudio en principio en relación a un tema genérico dado y la especificación de los tipos de estudio de que son susceptibles ciertas cuestiones o situaciones propuestas.
(Ver ejercicios 3, 4 y 5).

EJERCICIOS

2 (B)

Dados los temas de investigación que se señalan a continuación, y teniendo en cuenta las condiciones referidas en la introducción que se pueden considerar exigibles, por lo general, para que se pueda estimar aceptable un problema de investigación, se pide indicar razonadamente si los temas que comprende el enunciado de este ejercicio son o no adecuados para una investigación sociológica.

A) El ocio.

RESPUESTA

El ocio, sin más, no se puede considerar un tema adecuado por ser demasiado general y por no estar especificado en qué sentido se va a considerar el ocio (éste presenta muchas facetas, moral, religiosa, técnica, legislativa, etc., además de la sociológica), ni, supuesto que se va a enfocar su estudio sociológico, qué aspectos y qué problemas o cuestiones se pretenden investigar entre las muchas posibles que podrían ser objeto de consideración al respecto.

B) Deber de las empresas de realizar una política de relaciones humanas con sus empleados y obreros.

RESPUESTA

Tampoco es un tema aceptable, porque, aunque plantea un problema a resolver, éste no es de carácter sociológico, sino moral o político-práctico, según el punto de vista que se adopte, y porque en cuanto busca dilucidar lo que debe ser y no lo que es carece del enfoque empírico esencial en los estudios científicos.

C) El problema de si es mejor para el hombre el llamado estado de naturaleza o la vida en sociedad.

RESPUESTA

Este tema, además de plantear un problema de tipo valorativo, que implica un juicio de valor, no puede ser objeto de observación y experimentación, ya que hasta la fecha no poseemos ningún antecedente que no sea literario de la existencia generalizada de tal estado, supuesto que haya existido alguna vez en la realidad.

D) El estudio de las relaciones sociales de padres e hijos en una única familia española que vive en Laponia.

RESPUESTA

Es dudosamente admisible, porque no cumple la condición, dada su singularidad, de ser un estudio representativo de una situación o área social determinada.

E) El diferente nivel de rendimiento intelectual de los niños con padres y de los acogidos a hospicios.

RESPUESTA

No es un tema adecuado de investigación, por haber sido objeto de numerosos estudios e investigaciones, salvo que se abordaran puntos concretos aún sin aclarar o con un enfoque nuevo.

3

Teniendo en cuenta los aspectos y relaciones básicos que se pueden distinguir e los fenómenos sociales, enunciar algunas cuestiones que podrían ser objeto de investigación respecto a la institución social española del caciquismo.

RESPUESTA

Entre los aspectos básicos se señalan los siguientes:

Elementos, caracteres y notas de sus diversas manifestaciones en el espacio y en el tiempo.

Orígenes de esta institución y las razones y motivos sociales que explican su existencia.

Sus funciones y disfunciones respecto a la vida local española.

Las fuerzas sociales que han actuado con relación al caciquismo y las tensiones y conflictos sociales a que han dado lugar.

Valores y actitudes de la mentalidad caciquil.

Conducta moral y religiosa de los caciques.

Estructura social de las comunidades dominadas por el caciquismo.

Existencia de transformaciones a lo largo del tiempo en el caciquismo y su relación con el cambio social general.

Como relaciones se destacan las siguientes:

Relaciones recíprocas entre caciquismo y favoritismo.

Relación entre el caciquismo y el nivel económico y cultural de cada región.

Relación entre el caciquismo y los conflictos sociales.

Clases sociales y caciquismo.

Relación del caciquismo con la política, la religión, la educación y la economía.

Caciquismo y edad, sexo y regiones.

4

Siguiendo el procedimiento de los ejercicios anteriores, enumerar los aspectos e investigar que tendrían más aplicación y enunciar algunas de las cuestiones referentes a relaciones con otros fenómenos que podrían ser objeto de investigación respecto al abandono de familia.

RESPUESTA

Los aspectos generales que tienen más aplicación en este caso son los siguientes: descriptivo, funcional, evolutivo, conflictivo, dialéctico, histórico-social, prospectivo, o referente a posibles tendencias en el futuro del fenómeno, y motivos y consecuencias del abandono de familia.

Entre las cuestiones derivadas a relacionar el abandono de familia con otros aspectos sociales, se destacan los que se indican a continuación:

Su relación con la edad, el sexo, la clase social, el tipo de ocupación y el nivel de instrucción.

Su relación con la salud mental de los cónyuges.

Su relación con la existencia o no de hijos en el matrimonio.

Su relación con la situación económica.

Su relación con el tipo de comunidad y la región geográfica.

Su relación con las diferencias de edad conyugal y el ajuste sexual.

5 (A)

Supuesto el caso de coexistencia en una misma comunidad local de dos grupos étnicos; por ejemplo, en el País Vasco los aborígenes y los procedentes de otras regiones; se desea saber si se podrían estudiar comparativamente en ambos grupos los aspectos y relaciones indicados en la introducción y enunciar algunas cuestiones susceptibles, en principio, de investigación.

RESPUESTA

Los estudios comparativos de comunidades y fenómenos sociales es otro de los temas corrientes de investigación y una de las vías más fecundas de la investigación sociológica. Por ello, efectivamente es viable, en principio, el estudio en cada grupo étnico de los aspectos y relaciones básicos indicados en la introducción y su análisis comparativo.

Enunciado de cuestiones:

Actitudes, relaciones y distancia social entre los miembros de dichos grupos.

Relaciones entre la pertenencia a uno de los grupos y la clase social.

Relación entre la pertenencia étnica y la participación en el poder social, político y económico.

Estudio demográfico comparativo de cada grupo.

Conflictos sociales entre ambos grupos y motivos de ellos.

5 (B)

Existe un pueblo X, en Méjico, en el que desde hace varios años han convivido pacíficamente diferentes confesiones religiosas.

Otro pueblo Y, vecino y similar al anterior, ofrece, por el contrario, un espectáculo distinto. Las comunidades han tenido entre sí rencillas y aún luchas violentas. Determinar el problema que podría ser objeto de investigación en dichas comunidades según las indicadas situaciones de hecho.

RESPUESTA

Es un caso típico de hallazgo de un problema basado en la observación comparativa de realidades sociales contrapuestas. Del enunciado se deriva claramente que la cuestión a investigar serían los motivos de las diferencias de comportamiento en las relaciones entre los grupos de distinta confesión religiosa de las dos comunidades.

(Basada en F. Pardiñas, 1969. Metodología y técnicas de investigación en ciencias sociales. Siglo XXI, editores, p. 44).

6

Es un hecho el rápido desarrollo económico de España a partir del año 1959, especialmente.

También es un hecho, en España, desde el siglo pasado, el nivel económico notoriamente inferior de las regiones del Noroeste, Centro y Sur de la Península respecto a las del Norte, Nordeste y Levante.

¿Qué cuestión podría ser investigada habida cuenta de estos hechos?

RESPUESTA

Tenemos aquí una situación de partida: el desnivel económico entre las dos partes territoriales que pertenecen a una misma nación y una causa común: el proceso de desarrollo planificado producido a partir del año 1959.

Por ello, la cuestión a investigar sería, en principio, el efecto del desarrollo en ambas Españas: la incidencia o los efectos comparativos en ellas de la evolución económica producida, con objeto de ver si esta evolución ha acentuado o disminuido la divergencia en el nivel económico inicial.

5 . FORMULACION DE HIPOTESIS

NOCION E IMPORTANCIA

Etimológicamente, la palabra hipótesis tiene su origen en los términos griegos *thesis*, que significa lo que se pone, e *hipo*, partícula que equivale a debajo.

Hipótesis, es, pues, según sus propios términos, lo que se pone debajo, o se supone.

Etimológicamente, por tanto, las hipótesis no son otra cosa que suposiciones. De acuerdo con esta noción, las hipótesis presentan una gran generalidad. Se puede decir que casi todos los enunciados, afirmaciones o negaciones que utilizamos en el lenguaje oral y en el escrito se pueden considerar, por lo menos en algún aspecto, hipótesis.

Las hipótesis científicas también son, en general, suposiciones. Desde este punto de vista, son enunciados teóricos supuestos, no verificados pero probables, referentes a variables o a relación entre variables.

También científicamente, pero desde el punto de vista del problema a investigar, las hipótesis se pueden definir como soluciones probables, previamente seleccionadas, al problema planteado, que el científico propone para ver, a través de todo el proceso de la investigación, si son confirmadas por los hechos.

En esta definición se puede observar la estrecha relación entre la determinación del problema a investigar y las hipótesis. Sin embargo, hay que hacer la salvedad que las hipótesis se refieren especialmente, sobre todo, no a las investigaciones puramente descriptivas, sino a las investigaciones propiamente científicas o explicativas que intentan hallar la razón o el motivo de los problemas planteados.

Con esta salvedad se puede decir que la formulación de hipótesis –puede ser más de una, alternativas o no– es la operación subsiguiente a la determinación del problema en el proceso investigador.

En efecto, la especificación del problema a investigar no anticipa nada, como es obvio, sobre su solución, pues en otro caso dejaría de ser problema. Pero ocurre que, planteada una cuestión, se abren en la realidad ante el investigador muchas posibles respuestas, entre las cuales ignora, por supuesto, cuál es la verdadera.

Dado que el proceso de investigación científica explicativa no consiste en otra cosa sino en ver si una determinada respuesta a la cuestión planteada se ajusta o no a la realidad o es o no confirmada por los hechos, emprender sin más, una vez determinado el problema, la búsqueda de su solución, llevaría al investigador a actuar a ciegas y sin orden, al no saber en qué sentido orientar el hallazgo de la solución pretendida.

Por ello, el científico se ve obligado a plantearse un nuevo interrogante y a efectuar una nueva elección. Después de haberse preguntado qué investigar y buscando qué en la determinación del problema, debe inquirir ahora ¿cuál es la solución o soluciones probables a la cuestión planteada? En contestación a esta pregunta ha de efectuar, entre las diversas explicaciones posibles del fenómeno que se le ocurran, la elección de aquélla o aquéllas que le parezcan más plausibles o verosímiles, a fin de proceder a la comprobación en la investigación de su validez. Esta explicación o solución posible elegida no es otra cosa que la hipótesis según la definición expuesta.

Las hipótesis representan, pues, otra nueva concreción del tema a investigar, necesaria para proceder con el debido orden en la comprobación científica. Precisan, en general, el objetivo de ésta y orientan la búsqueda de explicación al problema planteado, lo que equivale a decir que iluminan el proceso de la investigación en sus distintas fases.

Por todo ello, las hipótesis son básicas en toda investigación científica. En la investigación social también lo son. Lo que pasa en ésta es que, debido a la complejidad de los fenómenos sociales, es sin duda más difícil que en las ciencias físicas y naturales, el hallazgo y la formulación de hipótesis significativas, como por lo demás, según hemos visto es más difícil la investigación en general.

Sin embargo, de lo anterior no se deriva que las hipótesis sean menos necesarias en la investigación social, sobre todo si se piensa que en toda acción de la investigación, planteada de un modo inteligente, con un sentido y fin determinado, late siempre una hipótesis, por lo menos en cuanto implica suponer que dicho planteamiento es mejor y más conveniente para llevar a cabo la investigación.

De las hipótesis se derivan las variables en estudio y se deben fundar en ellas la determinación del campo de investigación, de las informaciones a recoger, de los métodos a emplear y, en el curso de la investigación, de los datos o hechos válidos o de interés para el estudio.

NATURALEZA DE LAS HIPOTESIS: LOS ENUNCIADOS

Según la definición de hipótesis expuesta, las hipótesis son enunciados especiales caracterizados, desde el punto de vista epistemológico, por ser ideas supuestas no verificadas pero probables y, en cuanto a su contenido ontológico, por referirse a variables o relaciones entre variables.

Las hipótesis son, pues, ante todo y en primer lugar, enunciados que expresan lingüísticamente juicios, es decir, afirmaciones o negaciones sobre la realidad.

Los enunciados, que antes recibían el nombre de proposiciones, están formados, como elementos fundamentales, por los signos verbales que comprenden y por las ideas que expresan los signos verbales, las cuales recibían antes el nombre particular de juicios y hoy se llaman preferentemente proposiciones.

Por tanto se distinguen en el enunciado, como elementos fundamentales, su contenido o significado, lo que afirman o niegan, y su forma lingüística. Asimismo, su significado les viene de que enuncian o declaran algo de una cosa, es decir, de una realidad distinta e independiente

En consecuencia, son tres los niveles que se pueden distinguir en los enunciados: real, conceptual y terminológico.

Según el primero los seres o cosas no solamente existen, sino que existen de una manera determinada, tienen o no ciertas características

De acuerdo con el segundo, no sólo aprehendemos mediante el conocimiento las cosas en su conjunto, sino también en sus atributos, lo que nos permite formar intelectualmente juicios o proposiciones en que se afirmen o nieguen conceptualmente dichos atributos de las cosas.

Por último, conforme al tercero, los juicios o proposiciones formados precisan para su expresión o manifestación de una forma lingüística.

Entre el nivel lingüístico y el nivel conceptual hay una relación de designación, el enunciado designa un juicio, y entre el juicio o proposición y la realidad, una relación de referencia, el juicio se refiere a un determinado fenómeno o realidad.

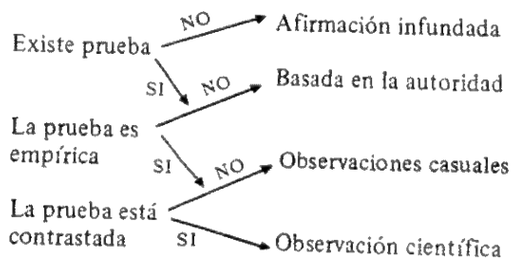
Estos tres niveles no se identifican. Los juicios o proposiciones pueden ser más o menos objetivos, representan mejor o peor la realidad, se puede expresar el mismo juicio por distintas palabras, por ejemplo de lenguas diferentes.

Una división importante de los enunciados, o afirmaciones y negaciones; desde el punto de vista científico, es la que los distinguen según se refieran o no a hechos verificables empíricamente.

Los primeros son los que se pueden comprobar mediante los sentidos o la observación empírica. En los segundos tal comprobación no es factible directamente porque comprenden términos morales, metafísicos o abstractos. Los enunciados o afirmaciones de hecho se pueden subdividir en infundados, que tienen lugar cuando no se aduce ninguna prueba que los apoye; por ej., si se dice: existen elefantes sin trompa; de autoridad, cuando esta prueba se basa solamente en la opinión de un experto o científico relevante en el campo de que se trate; casuales, cuando existe la prueba y ésta es de carácter empírico, pero es aislada y no contrastada científicamente, por ej.: la de los que dicen haber visto OVNIS y, por último, científicos, cuando se apoyan en una prueba empírica comprobada científicamente.

Esta clasificación es importante porque puede servir para juzgar, la validez, desde un punto científico, de las afirmaciones científicas que oigamos o leamos.

Según Paul C. Stern (1979) se puede formar este esquema:



ESTRUCTURA DE LAS HIPOTESIS

De conformidad con la naturaleza de los enunciados y sus niveles o aspectos, la estructura de las hipótesis se puede considerar desde un cuádruple punto de vista: científico, gramatical, lógico y matemático. Desde un punto de vista científico, los elementos estructurales de las hipótesis son las unidades de observación, las variables y las relaciones que unen a ambas y las variables entre sí.

Las unidades de observación son las personas, grupos, objetos, actividades, instituciones y acontecimientos sobre los que versa la investigación social. Esta estudia, respecto a las unidades de observación, determinadas modalidades y características, bien cualitativas, bien cuantitativas, que constituyen las variables.

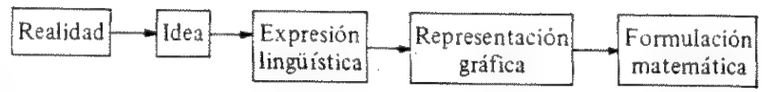
En términos gramaticales las hipótesis adoptan la forma de oraciones de tal carácter, en las que las unidades de observación son los sujetos de dichas oraciones, y las variables, los atributos.

Desde un punto de vista lógico, una hipótesis es un enunciado o sentencia, considerando como tal (Cfr. Sacristán, 1973, p. 29) la expresión lingüística de un juicio o proposición, y sus elementos son también las variables, los términos o categorías lógicas de unión, atribución y relación de las variables entre sí y con los nombres o constantes, que es como se llaman en la terminología lógica moderna, las unidades de observación.

Matemáticamente, la estructura de las hipótesis se manifiesta mediante funciones y ecuaciones en las que las variables y los coeficientes que las afectan se expresan mediante letras de los alfabetos griego y latino y los que hemos llamado términos lógicos por los signos indicativos de las operaciones y relaciones matemáticas.

Hay que señalar que las relaciones entre las variables que forman las hipótesis cuando son varias, se representan también gráficamente y, en este sentido, se puede hablar también de una estructura gráfica de las hipótesis.

La relación entre estos distintos puntos de vista estructurales de las hipótesis queda expresada en el siguiente esquema, que revela su íntima conexión:



La comprensión de la realidad da lugar a ideas o juicios; éstos se expresan lingüísticamente mediante enunciados; los enunciados se representan gráficamente mediante flechas que unen los signos de las variables entre sí; y, en fin, la representación gráfica se puede expresar matemáticamente mediante funciones y ecuaciones.

El esquema muestra a su vez cómo, en el proceso del conocimiento en que consiste la investigación científica, se pasa de la realidad a la idea por el pensamiento; del pensamiento al lenguaje por la expresión oral o escrita de las ideas; del lenguaje a la representación gráfica (modelo gráfico) mediante flechas que relacionan las variables de acuerdo con las relaciones que existan entre ellas según las hipótesis; y de la representación gráfica a la expresión matemática, formulando ecuaciones (modelo matemático) que especifiquen matemáticamente las relaciones entre las variables del modelo gráfico. (Ver ejercicio 8.)

CONDICIONES DE LAS HIPOTESIS

Respecto a esta cuestión, es necesario en la práctica conocer tales condiciones y saber aplicarlas, si se quiere evitar que hipótesis que no reunan las condiciones debidas, den lugar a la mala orientación de todas las fases de una investigación y, en último término, al fracaso de ésta.

Las condiciones que deben reunir las hipótesis, las cuales, como en el caso de los requisitos del problema investigado, pueden ser utilizadas para ver si una hipótesis es o no correcta, como se hace en los ejercicios propuestos, son (Vid. Goode y Hatt, 1952) las siguientes:

a) Las hipótesis deben ser conceptualmente claras y fácilmente comprensibles. Esto implica no sólo que no deben contener términos rebuscados y artificiosos, ni siquiera, a ser posible, excesivamente técnicos, de forma que sean accesibles al mayor número de personas, sino también que los conceptos empleados deben utilizarse en un sentido riguroso y preciso, definido previamente con el mayor cuidado, de forma que se excluya toda ambigüedad.

b) Los términos usados deben poseer una realidad empírica. Esta condición prohíbe el uso de términos morales y trascendentes y exige que se puedan observar empíricamente las cualidades o significados que denotan.

c) Las hipótesis deben ser susceptibles de verificación mediante el empleo de técnicas asequibles. Esta condición es una consecuencia de la anterior. Que los términos empleados posean un alcance empírico lleva consigo la posibilidad de verificación de las hipótesis, si existen técnicas adecuadas. Pide, pues, en último término, que se tenga en cuenta al formular las hipótesis si existen técnicas para contrastar con la realidad las variables que las forman.

d) Las hipótesis deben ser específicas o capaces de especificación. Según este presupuesto, se requiere que, sobre todo si las hipótesis son demasiado amplias y generales, y no se pueden verificar directamente, se pueda concretar su sentido y desarrollarlo en subhipótesis aclaratorias.

e) Las hipótesis deben hallarse en conexión con las teorías precedentes. Según este criterio se debe procurar que las hipótesis elegidas estén vinculadas en forma de confirmación, precisión, revisión, etc., a logros teóricos anteriores. Los logros científicos tienen lugar generalmente mediante nuevas aportaciones que enriquecen, revisan, transforman o puntualizan anteriores adquisiciones.

f) Las hipótesis deben poseer un cierto alcance general. En todo caso, como ya se indicó al hablar en el punto anterior de la exigencia de que los problemas a investigar sean reprentativos, las hipótesis no deben referirse a uno o unos pocos hechos, sino que deben presentar un valor de generalidad en el campo o sector al que se refieren. Sólo así se podrán descubrir regularidades en los fenómenos y la investigación tendrá un valor científico.

g) Las hipótesis, en fin, deben ofrecer una respuesta probable al problema objeto de la investigación. Esta condición, por lo demás obvia, en cuanto se deriva inmediatamente de la misma definición de hipótesis, proscribe no sólo las hipótesis que no tengan nada que ver con el problema propuesto o no guarden relación con él, sino también aquellas que, aunque tengan esta relación, sea muy poco verosímil que la investigación a realizar confirme que en realidad representan la solución buscada. (Ver ejercicio 7)

TIPOS DE HIPOTESIS

Las hipótesis se pueden clasificar atendiendo a su:

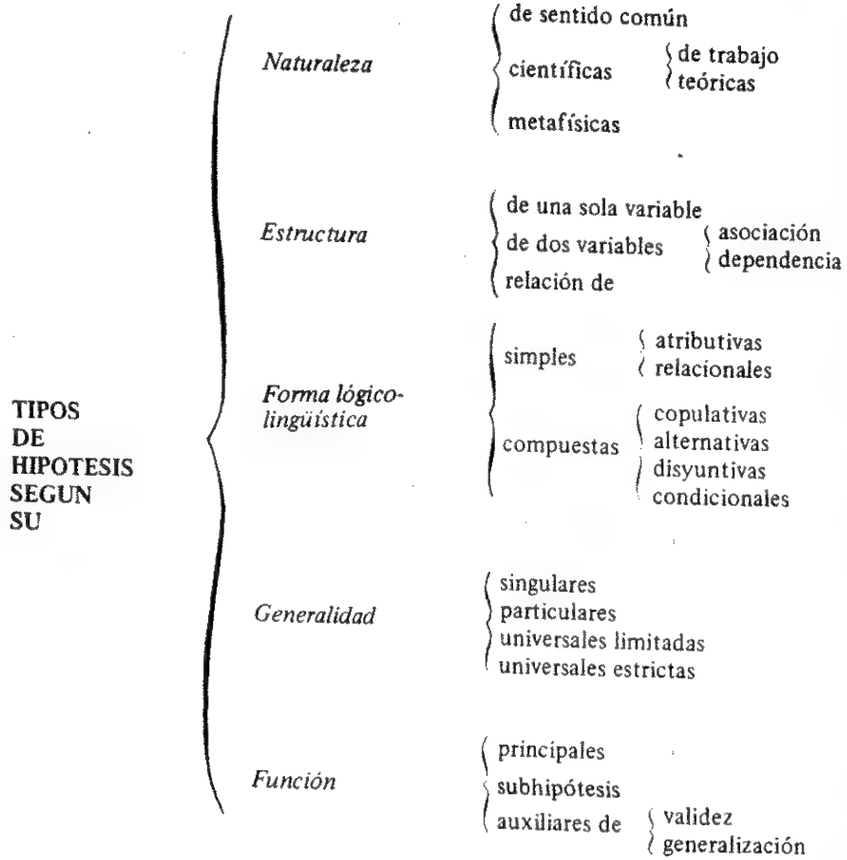
1. Naturaleza.
2. Estructura.
3. Forma lógico-lingüística.
4. Generalidad, y
5. Función en la investigación.

1. *Según su naturaleza*, las hipótesis se pueden distinguir, siguiendo a Wartofski (1973, Vol. I, p. 249) en hipótesis de sentido común, hipótesis científicas e hipótesis metafísicas.

Las hipótesis de sentido común son suposiciones sobre hechos y sucesos de la vida que todo el mundo puede verificar de una manera simple y directa mediante sus sentidos. Por ejemplo, una hipótesis de este tipo sería la siguiente: Si ponemos al fuego un cazo de líquido hervirá pasados unos minutos.

Las hipótesis científicas son aquellas suposiciones formadas por términos científicos y cuya verificación exige la realización de algún experimento o investigación.

Por último, como hipótesis metafísicas se pueden considerar bien las conjeturas cosmológicas, o suposiciones sobre la realidad totalmente ge-



nerales, como el universo está escrito en lenguaje matemático, o bien las que se refieren a enunciados que trascienden la experiencia científica, por lo menos hasta el presente. Por ejemplo, el alma es inmaterial.

Las hipótesis científicas se pueden subdividir en hipótesis de trabajo y teóricas.

Las primeras son las hipótesis específicas de una investigación determinada, es decir, las que en su formulación están ya concretadas con referencia a una investigación, a un tiempo, a un lugar y a una población definidos.

En último término, las hipótesis de trabajo deben establecer relaciones entre variables empíricas, inmediatamente observables, o indicadores.

Las hipótesis que no tengan esta concreción son las que se pueden llamar abstractas. Dentro de esta categoría cabe todo enunciado teórico, esté o no contrastado con la realidad. Si lo está porque pueden constituir nuevamente hipótesis de investigación con relación a poblaciones y tiempos no investigados. Si no lo está, con mucha mayor razón dada la necesidad de su verificación empírica.

2. *De acuerdo con su estructura,* las hipótesis se pueden referir a una sola variable, a dos o más. A su vez, en el caso de referirse a más de dos, pueden establecer una relación de asociación simple entre las variables en cuestión o una relación de dependencia. Por tanto, es posible distinguir los siguientes tipos de hipótesis:

- Hipótesis con una variable.
- Hipótesis con dos o más variables y relación de asociación.
- Hipótesis con dos o más variables y relación de dependencia.

El primer tipo, las hipótesis de una sola variable, se caracterizan por postular la existencia de determinadas uniformidades o regularidades empíricas en una población o universo. Las uniformidades postuladas pueden ser de carácter cualitativo o cuantitativo, temporal, espacial, etcétera. Incluso pueden describir un proceso o producción sucesiva de una circunstancia en la población. Estas son hipótesis simplemente descriptivas.

El segundo tipo, las hipótesis con dos o más variables, establece una relación simple de asociación o covarianza entre ellas, es decir, de cambio en una variable acompañado de un cambio correlativo en la otra del mismo o distinto signo, sin que se precise si una variable influye en el cambio de la otra. Entre este tipo de hipótesis se deben encuadrar todas aquellas que implican una relación entre las variables que no sea de causalidad en el sentido de influencia; por ejemplo, de reciprocidad, igualdad, superioridad, inferioridad, precedencia, sucesión, etc.

Las hipótesis con dos o más variables y relación de dependencia son las más interesantes desde un punto de vista científico, pues la ciencia busca explicar o conocer las razones o motivos de los fenómenos. Son las hipótesis propiamente explicativas, las que postulan los motivos, causas o efectos de los fenómenos.

3. *Teniendo en cuenta su forma lógico-lingüística,* las hipótesis pueden ser simples o compuestas, según estén formadas por un solo enunciado o por dos o más.

Las simples se pueden dividir, a su vez, en atributivas que implican la asignación a un sujeto de una característica, atributo o variable y relacionales que relaciona entre sí diversos sujetos o variables.

Asimismo las compuestas, que implican el establecimiento de una relación entre los enunciados simples que comprenden, se dividen en copulativas, disyuntivas, alternativas y condicionales, según que el elemento de unión en la hipótesis de los distintos enunciados sea una conjunción copulativa, disyuntiva, alternativa o condicional, respectivamente.

4. *Conforme a su generalidad,* se pueden clasificar en singulares si se refieren a un solo individuo; particulares que atañen a sólo una parte de una población o conjunto; universales, limitadas, que aluden a toda una población, pero dentro de un espacio o tiempo definidos, y universales estrictas que no tienen límite alguno, ni en la población, ni en el tiempo ni en el espacio. (Ver ejercicio 20).

5. Finalmente, Mora y Araujo (1973, p. 9) distingue *por su función*, los siguientes tipos de hipótesis:

1. Hipótesis sustantivas que relacionan variables conceptuales o generales.
2. Hipótesis que relacionan los indicadores derivados lógicamente de las hipótesis sustantivas. Son las que se contrastan empíricamente.
3. Hipótesis auxiliar de validez que supone hay una relación entre los indicadores y las variables conceptuales o generales.
4. Hipótesis auxiliar de generalización de la muestra a la población.

Las hipótesis sustantivas son las iniciales, las que motivan la investigación.

Las segundas son hipótesis de segundo grado o subhipótesis en relación a las hipótesis sustantivas. Concretan éstas en relación a variables más cercanas a la realidad que las variables generales o conceptuales como dice Mora y Araujo.

Con la formulación de estas subhipótesis se realiza la especificación de las hipótesis. Esta susceptibilidad de especificación es, como vimos, una de las condiciones que deben reunir las hipótesis generales.

Las hipótesis auxiliares son necesarias, por ejemplo, en la operacionalización de variables. Son las que conectan entre sí conceptos de distinto nivel de abstracción. Estas hipótesis fundamentan lógicamente que el paso de un nivel a otro está justificado.

El cuarto tipo se refiere a las investigaciones hechas con muestras. En ellas se observa la muestra en lugar y representación de las poblaciones enteras.

Estas investigaciones, que son las corrientes en Sociología, implican en general y con referencia a las distintas relaciones entre variables detectadas en la investigación, la hipótesis de la validez de la generalización o de la extensión de los resultados obtenidos en la muestra a la población total.

La negación de esta hipótesis es lo que se llama la hipótesis nula que se contrasta con los llamados test estadísticos, como veremos en la parte de análisis estadístico. (Ver ejercicio 9)

FORMALIZACION LOGICA DE LAS HIPOTESIS

Consiste en la expresión de los enunciados que forman las hipótesis mediante los signos lógicos correspondientes.

En el caso de los enunciados simples atributivos, por ejemplo el hombre es mortal, teniendo en cuenta que las constantes nominales y predictivas se representan respectivamente por las primeras letras minúsculas y mayúsculas del abecedario latino, la fórmula lógica de este enunciado sería C (b), que se lee b es C. A efectos, pues, de formalización se considera que la cópula y el predicado forman una unidad.

Ejemplos de enunciados de relación son: Juana está casada con Pedro y Alcalá está entre Madrid y Barcelona. En ellos Juana y Pedro son los términos de la relación en el primero, Alcalá, Madrid y Guadalajara en el segundo y "estar entre" y "casados" la relación existente entre ellos.

Designando Juana y Pedro, como constantes, por las letras minúsculas a y b y Alcalá, Madrid y Guadalajara por c, d y e. La fórmula lógica de estos enunciados sería Rab o R(ab) y Rcde o R(cde), que se puede leer, relación entre a y b, en el primer caso, y entre c y d y e en el segundo.

La formalización de las hipótesis compuestas se realiza designando cada enunciado simple bien por su fórmula propia, bien mediante las letras minúscula p, q, r... y empleando para representar su unión en el enunciado compuesto las llamadas conectivas siguientes:

- La condicional: si... entonces, o implicación simple que se representa por $p \rightarrow q$ (si p, entonces q.)
- La conjuntiva, que significa la afirmación simultáneamente de dos enunciados y su signo es $\wedge$.

- La disyuntiva, que los afirma alternativamente, en cuyo caso su signo es v, o de forma recíprocamente excluyente de ambos enunciados entre sí, con signo entonces de w, y.
- La implicación doble o equivalencia que se representa por la doble flecha ↔ o por ≡, signos que significan, si y sólo si.
- Por último, la cantidad de los enunciados: todos, alguno, se formaliza mediante el cuantificador universal, que se representa en el lenguaje lógico por (x) y ∀ y quiere decir "para todo x" y el existencial ∃ que significa "hay algún X".

DETERMINACION DE HIPOTESIS

En el proceso seguido para la determinación de hipótesis se pueden distinguir dos pasos principales: a) hallar el aspecto, relación o factor que pueda ser solución a la cuestión investigada, y b) expresar dicho aspecto, relación o factor en forma adecuada susceptible de verificación.

A su vez, en el primer paso se pueden distinguir dos casos, según que el problema a investigar haya sido o no determinado de antemano, ya que se puede dar el caso, y de hecho se da en la realidad, de tener la intuición de una hipótesis sin existir una formulación explícita previa del problema. En este caso de la hipótesis se pasa a la especificación del problema en lugar de ser al revés.

Tanto en un caso como en otro, no existen reglas previas que resuelvan la cuestión de hallar el aspecto, relación o factor clave base de la hipótesis. También aquí, como en la determinación del problema, nos encontramos en el dominio de la intuición, agudeza y penetración del investigador.

No obstante, si el problema ha sido determinado de antemano, caso éste el más normal, la tarea es más fácil, pues resulta limitada por la formulación del problema a investigar al que debe dar respuesta.

En todo caso, las fuentes principales de hallazgo de hipótesis son, en primer lugar, el conocimiento vivo y experimental de las realidades sociales, el dominio crítico de la teoría de la disciplina en cuestión, con atención especial a sus imperfecciones y lagunas, la cultura científica general, en cuanto muchas hipótesis se basan en la aplicación analógica al campo social de hallazgos y descubrimientos en otras áreas y, acaso de modo especial, la preocupación científica y la vivencia de los problemas investigados.

Aun cuando el hallazgo de hipótesis sea obra fundamentalmente imaginativa, existen técnicas que pueden ayudar a la imaginación en esta tarea y, por otra parte, es factible incluso determinar hipótesis mediante procedimientos más o menos lógicos, entre los que cabe destacar la inducción y la deducción.

Con relación a la imaginación, C. Javeau en su obra *Enquête par questionnaire* (1971, p. 11), propone concretamente el uso de las siguientes técnicas para promover el hallazgo de hipótesis:

1. El examen de la literatura sobre la materia para deducir *temas de referencia*, que sirvan para el enunciado de hipótesis de trabajo.
2. La *reunión-discusión* del grupo de investigación al efecto, o *brainstorming*. Las ideas surgidas se pueden someter a la técnica de análisis de contenido.
3. La entrevista de grupo con personas competentes o interesadas en la investigación.
4. Entrevistas en profundidad a personas seleccionadas por su información o implicación en la encuesta.

Mediante la aplicación de la imaginación y en algún modo también de la deducción, se pueden hallar hipótesis buscando las causas de efectos determinados o, al revés, los efectos de ciertas causas. Partiendo, como explica Zettemberg, de una variable específica como determinante o variable independiente, se pueden formular listas de enunciados formados por dichas variables y los diferentes resultados o variables dependientes a que pueden dar lugar. (Ver ejercicios 16 y 17)

Zettemberg muestra igualmente la formación de hipótesis, ésta ya deductiva totalmente, sobre relaciones entre variables mediante la combinación entre sí de elementos comunes de otros enunciados o la sustitución de los términos que los forman por otros equivalentes (Ver ejercicios 14, 15 y 18)

Otra forma de inferir hipótesis que se puede considerar deductiva es las deducciones de principios admitidos. De un principio admitido se pueden deducir consecuencias que si se expresan en términos observables pueden constituir hipótesis a efectos de su comprobación empírica.

Por ejemplo, si damos por admitido el principio de que los españoles son individualistas, de este principio podemos derivar consecuencias que podrían formar hipótesis a efectos su investigación empírica. Por ejemplo, una de las conclusiones podría ser la siguiente, si los españoles son individualistas, entonces la tasa de asociaciones voluntarias de diverso

tipo será inferior a la de otros países. Este tipo de raciocinio, que adopta la forma de "si... entonces" y que de principios más generales deduce consecuencias directamente observables, tienen gran importancia científica no sólo como fuente de hipótesis concretas de trabajo sino como medio de prueba de dichos principios generales. (Ver ejercicio 21)

Analógicamente, se pueden construir hipótesis mediante argumentos de analogía. Bunge distingue dos tipos de salto analógico:

1. *Analogía sustantiva:* Cuando la respuesta de una organización a un estímulo o fenómeno sugiere la hipótesis de que en otra organización diferente tendrá lugar la misma relación estímulo-respuesta; o bien, cuando se trasladan las propiedades de un objeto a otro: suponer que los seguidores de líderes autoritarios lo son también.

2. *Analogía estructural:* Si se sospecha que la ley de crecimiento de una población tendrá la misma forma que la ley del crecimiento de un individuo, o que las leyes de aprendizaje de los animales superiores valen también para los hombres.

Inductivamente, se puede construir igualmente hipótesis sobre la base del examen de casos diversos. Se distinguen dos tipos de generalización inductiva:

1. La inducción de primer grado, o inferencia, que va de enunciados particulares a enunciados generales, v.g. si del examen de ciertos números de casos individuales se infiere que "el estudio del francés interfiere en el estudio del italiano".

2. Inducción de 2° grado, o generalización de generalizaciones de primer grado. Ejemplo si de la generalización anterior se induce que el aprendizaje de cualquier tema interfiere en el de cualquier otro tema contiguo.

Otro tipo de formación de hipótesis relacionada con la inducción, es la generalización empírico-inductiva. Por ejemplo, si previa la realización de investigaciones en diferentes poblaciones se hubiera llegado a la conclusión de que las personas de clase media baja presentan en elevada proporción un alto grado de autoritarismo, esta generalización empírica podría ser una hipótesis en relación a su investigación en una población concreta no investigada al respecto. (Ver Ejercicios 10 a 13 y 19)

EJERCICIOS

7

Dadas las siguientes hipótesis:

1. Los que hacen bien al prójimo se hallan en gracia de Dios.

2. Los malos profesores deben ser expulsados, porque son el motivo de que la marcha de la Universidad sea deplorable.

3. Las personas de prestigio se convierten en asociados, cuya interacción es buscada, y tienden así a obtener posiciones centrales en el grupo. (Zettemberg, 1971, pág. 78.)

4. En la sociedad de los marcianos no existen desigualdades sociales.

5. La tipología idiosincrática del campesino es un producto irracional de la acción combinada de amalgamas sociales y culturales.

6. Cuanto mayor sea el índice de urbanización de una región tanto menor será su tasa de fertilidad.

7. "La delincuencia juvenil aumenta en los grupos de emigrantes", y "la delincuencia de los drogadictos es una consecuencia de que llevan los cabellos largos", formuladas en un estudio sobre "los drogadictos y la delincuencia juvenil".

Y teniendo en cuenta las condiciones de las hipótesis señaladas en la introducción, determinar razonadamente si las hipótesis que comprende el enunciado de este ejercicio son o no adecuadas.

RESPUESTA

1 y 4. No son admisibles estas hipótesis, porque no puede ser objeto de verificación empírica si uno se halla o no en estado de gracia, ni tampoco, hasta ahora, si en la sociedad de los marcianos, supuesto que exista, se dan o no desigualdades sociales.

2. Tampoco es adecuada porque postula una decisión moral y comprende en las variables dos juicios de valor, malos profesores y marcha deplorable de la universidad, cuyo sentido depende de lo que cada uno entienda por *malo* y *deplorable*, según su sistema de valores.

3. Se trata de una hipótesis aceptable porque reúne las condiciones indicadas de ser conceptualmente clara, estar formada por términos empíricos, ser verificable y estar relacionada con la teoría de la estratificación.

5. Esta hipótesis debe ser rechazada porque no es conceptualmente clara, sus términos son rebuscados y artificiosos y no es fácilmente comprensible.

6. Esta hipótesis sí es válida. Es conceptualmente clara y sencilla. Sus términos tienen un alcance empírico (existen índices para medir el grado de urbanización). Puede, en fin, ser objeto de verificación, es general y está relacionada con la teoría correspondiente. Aunque es general, puede descomponerse en otras hipótesis más concretas, como por ejemplo:

a) Con el aumento de la urbanización la tasa de fertilidad descenderá menos en las ciudades de 10.000 a 100.000 habitantes que en las de más de 100.000.

b) El descenso de la tasa de fertilidad consecuencia del desarrollo de la urbanización será menor en la clase obrera que en las profesiones liberales, empleados y técnicos.

c) El descenso de la fertilidad consecuencia de la urbanización será debido a una elevación en la edad de contraer matrimonio, o al aumento de la proporción de matrimonios sin hijos o con un hijo o dos; o, en fin, al aumento proporcional del número de solteros, etc.

7. Tampoco son válidas estas hipótesis. La primera porque no ofrece una respuesta ni guarda relación directa con el problema que se dice en el enunciado objeto de la investigación y la segunda porque es muy poco probable que los cabellos largos puedan ser en la realidad la causa de la delincuencia de los drogadictos.

8

Dadas las hipótesis (A-F) de las páginas siguientes:

1. Analizar su estructura.

2. Escribir su fórmula lógica.

3. Indicar los tipos a los que pertenecen por este orden:

a) Naturaleza
b) Estructura
c) Forma lógico-lingüística y
d) Generalidad

A) Si hay un aumento en el número de miembros de un grupo completamente no estructurado, siempre hay un aumento en el anonimato de las acciones del grupo.

RESPUESTA

Unidades de observación: Grupos no estructurados.

Variables: Número de miembros y carácter anónimo de las acciones del grupo.

Términos de relación: "Si hay... siempre hay", que expresan una relación de dependencia suficiente de la segunda variable respecto de la primera, pues basta con que se dé la primera para que siempre tenga lugar la segunda.

Fórmula lógica: $(x)\,(Ax) \rightarrow (Bx)$

Que se lee: para todo grupo completamente no estructurado, x, si x es A: aumenta en el número de sus miembros; entonces x es B: aumenta el anonimato de las acciones del grupo.

Tipos:

- Científica, porque su verificación exigiría realizar una investigación;
- De dos variables en relación de dependencia;
- Compuesta condicional. "Si hay..." condición "siempre hay...", resultado condicionado;
- Particular, se refiere no a todo grupo sino a grupos de características particulares.

B) Todas las culturas ganaderas son nómadas.

RESPUESTA

Estructura:

Unidades de observación: las culturas ganaderas.

Variables: Una sola, nómadas.

Términos de relación: "Todas... son", que indica una relación de atribución cualitativa general de la variable a la unidad de observación.

Fórmula lógica: $(x)\,[(Cx) \rightarrow (Dx)]$

Que se lee: para todo x, si es una cultura ganadera, entonces, x es nómada.

Tipos:

- Científica, por la misma razón que la anterior;
  De una sola variable;
- Simple atributiva, y
  Particular, por la misma razón que antes.

E "Si el costo de evitar la interacción es bajo y si existen fuentes de satisfacción alternativas, a mayor frecuencia de interacción, mayor simpatía mutua."

RESPUESTA

Unidad de observación: No se encuentra explícita. Ya que la unidad de interacción son las personas, hay que entender que ellas son la unidad de observación.

Variables: Son cuatro. Primera, costo de evitar la interacción. Segunda, fuentes de satisfacción alternativa. Tercera, frecuencia de la interacción. Cuarta, simpatía mutua.

Términos de relación: Son "si...si...a mayor... mayor", que indican una relación de covariación directa de las variables tercera y cuarta condicionada a la presencia de las otras dos variables.

Fórmula lógica: $(x)\,[(Ax) \wedge (Bx)] \rightarrow (Cx) \rightarrow (Dx)$

Que se lee: Para todo x, si x es A, costo bajo de evitar la interacción, y x es B, existencia de fuentes de satisfacción alternativas, entonces si x es C, mayor frecuencia de interacción, entonces x es D, mayor simpatía mutua. Si se llama a los tres enunciados, por su orden, p, q y r, entonces la fórmula sería $(p \wedge q) \rightarrow r$.

Tipos:

- Científica.
- De cuatro variables, las dos últimas de covariación condicionada.
- Universal, se refiere a todo grupo.
- Compuesta condicional.

D) En España sólo alrededor del 5 por 100 de los estudiantes universitarios proceden de la clase obrera.

RESPUESTA

Unidad de observación: España.

Variables: Sólo existe una. Procedencia de la clase obrera de los universitarios españoles.

Término de relación: Es "sólo alrededor", que indica una atribución cuantitativa aproximada de la variable a la unidad de observación.

Fórmula lógica: $\exists$ x (Px) $\wedge$ (Qx)

Que se lee: Hay algún x, España, tal que x es España y x tiene alrededor del 5 por 100 de estudiantes de clase obrera.

Tipos: (Se invita al lector a señalarlos por sí solo.)

E) La formación social y cultural de los jueces italianos tiende a hacerles extraños a la sociedad en la que están llamados a juzgar, condicionando así el ejercicio de su función social. (De A. Pagoni. El perfil ideológico del juez. Quaderni di Sociología 20, 1, 1972.)

RESPUESTA

Unidad de observación: los jueces italianos.

Variables: Cuatro. La formación social; la formación cultural; extraños a la sociedad, y ejercicio de su función social.

Términos de relación: Y, que une la formación social y la cultural, tiende a hacerles, que indica una relación de influencia y condicionando, que la indica de determinación.

Fórmula lógica: (x) [(Px) $\wedge$ (Qx) $\rightarrow$ (Cx) $\wedge$ (Dx)]

Que se lee para todo x, juez italiano, si x es P y Q, formación social y cultural, entonces x es C, extraños a la sociedad y D, con un ejercicio de su función social condicionado.

Tipos:

- Científica de trabajo, se refiere a una investigación determinada.
- De cuatro variables en relación de dependencia.
- Universal, restringida espacialmente a Italia.
- Compuesta, condicional y conjuntiva.

F) Los esquemas de comportamiento innatos son más estables que los adquiridos (Bunge, 1972, 359).

RESPUESTA

Unidad de observación: Todo individuo, incluso organismo.

Variables: Dos. Estabilidad del comportamiento innato y estabilidad del comportamiento adquirido.

Términos de relación: más que, que indica una relación de superioridad en la estabilidad de un comportamiento respecto al otro.

Fórmula lógica: Establece una relación, "más estable que" entre dos variables, "los esquemas de comportamiento innato" y "los adquiridos". Implícitamente ambas se refieren a un organismo. Sin embargo a efecto de su representación lógica se puede ver que implica un enunciado condicional, representado por la fórmula lógica siguiente:

$$(x)\,(y)\,(z)\,[\,O\,(x)\,\&\,I\,(y)\,\&\,A\,(z) \rightarrow R\,(yz)\,]$$

Que se lee: Para todo x, para todo y, y para todo z, si x es O, un organismo, si y es I, un comportamiento innato, y si z es A, un comportamiento adquirido, entonces y, z están en la relación R, o sea el primero es más estable que el segundo.

Tipos:

- Científica, teórica.
- De dos variables en relación de asociación.
- Singular relacional.
- Universal.

9

Dadas las siguientes hipótesis, indicar su tipo según su función

A) El consumo de energía por habitante es un indicador que representa válidamente la variable general, nivel económico de un país.

RESPUESTA

Es una hipótesis auxiliar de validez que se sobreentiende en toda elección de indicadores de una variable general.

B) Los resultados obtenidos en una muestra, según la cual un 35 por 100 prefiere un sistema económico socialista autogestionario, son válidos y significativos en relación a la población total de que procede la muestra.

RESPUESTA

Es una hipótesis auxiliar de validez de la generalización de los resultados de la muestra a la población entera.

10

Supuesto que se quieren investigar los motivos sociales de la separación familiar, y teniendo en cuenta que los vínculos sociales y familiares suelen ser más fuertes en las comunidades pequeñas que en las grandes, en las cerradas que en las abiertas y en los estratos de costumbres tradicionales que en los de formas de vida más libres, ¿qué hipótesis se podrían formular respecto al problema indicado?

RESPUESTA

Según los datos anteriores se puede presuponer que la estabilidad familiar depende, sociológicamente hablando, por una parte, de la presión del medio social, y por otra, de la libertad de costumbres de cada estrato social, y que, por tanto, debe existir una relación de dependencia de la variable separación familiar respecto al tipo de área ecológica en que se vive, rural o urbana, y en ésta según sea la comunidad más o menos abierta o cerrada, así como con la mayor o menor libertad de costumbres del estrato social al que pertenecen las familias.

De aquí que sea admisible formular, en consecuencia, las siguientes hipótesis:

1. La ruptura familiar será menos frecuente, en igualdad de circunstancias, en las comunidades rurales que en las urbanas.

2. Tanto menos frecuente será la separación matrimonial cuanto los lazos sociales de las comunidades urbanas sean más cerrados.

3. Tanto más frecuente será la separación familia cuanto más liberales sean las costumbres del estrato socio-económico a que pertenecen las familias.

(Basadas en el trabajo de Trevor Noble. Family Breakdown and social Network. The British Journal of Sociology, 21, 2, 1970.)

10 B

Cuestión a investigar, "actitudes de la clase obrera ante los conflictos universitarios". Son de todos conocidos los frecuentes y agudos disturbios universitarios en el mundo y concretamente en España en la década del 70. Pero se han realizado escasas investigaciones para conocer la reacción y actitudes de los demás grupos sociales ante esta situación. Una encuesta realizada en U.S.A. (H. E. Ransdorf. Blue collar anger: reaction to student and black protest. American Sociological Review, 37, 31, 1972) confirmó la hipótesis de que se daba un mayor antagonismo y hostilidad hacia la conducta de los estudiantes en los ambientes de la clase trabajadores que en el de los empleados y profesionales. ¿Qué hipótesis cabría formular con referencia a un estudio empírico a realizar en España sobre esta cuestión?

RESPUESTA

Teniendo en cuenta que la instrucción universitaria española hasta ahora es de modo predominante de tipo clasista, dada la mínima participación en ella de estudiantes de procedencia obrera, la información acaso más deficiente de este sector sobre los conflictos en cuestión, el antagonismo de clases y la tendencia observada en los obreros a adoptar actitudes más rígidas e intolerantes, parece que se podría mantener la indicada hipótesis base de estudio realizado en U.S.A. y establecer como punto de partida de la investigación a contrastar en la realidad, que

La actitud de antagonismo hacia los estudiantes universitarios contestatarios mayor relativamente entre la clase trabajadora que entre los empleados y profesionales.

11

Se ha hablado y escrito bastante de la emergencia en los últimos años de una nueva clase obrera, formada por obreros calificados o semicalificados y con ingresos relativamente elevados. Igualmente se ha observado que la autodesignación de la clase a que uno pertenece sólo puede tener valor indicativo con referencia a la concepción de la estructura de clases en el grupo encuestado.

Por ello representa una cuestión a investigar examinar cómo la nueva clase obrera, en el sentido indicado, concibe la estructura de clases y se sitúa en ella comparativamente respecto a los obreros "tradicionales" y empleados.

Formular razonadamente la hipótesis o hipótesis que se crean pertinentes respecto a esta cuestión a investigar.

RESPUESTA

Si se tiene en cuenta que el mayor nivel económico disminuye normalmente la tensión y oposición clasista y la mayor valoración por los trabajadores de los aspectos materiales y positivos de la vida social en relación a los de apariencia y prestigio, respecto a los empleados, se puede formular la siguiente hipótesis:

"Los obreros de la 'nueva clase' tienden a interpretar la estratificación social en términos de riquezas, y no de lucha y poder como los obreros 'tradicionales', ni tampoco de prestigio, como los empleados y profesionales, y se sitúan a sí mismos en la estructura de clases de acuerdo con este esquema."

Esta hipótesis puede ser objeto de verificación porque sus términos presentan una referencia empírica, son susceptibles a ser analizados por medio de indicadores adecuados y empleando las técnicas existentes y guarda relación con la teoría de las clases y la estructura social. (Basado en J.H. Goldthorpe. L'image des classes chez les travailleurs manuels aisés. Revue Française de Sociologie, 11, 3, 1970.)

12

Según los estudios de sociología de la familia en U.S.A. (Vid. Ricards Centers and Bertam H. Raven, Conyugal power structure: a reexamination. American Sociological Review, 36, 2, 1972), uno de los aspectos más importantes de la estructura de la familia hoy es el poder comparativo de decisión del marido y la mujer. Indica el papel que juegan en el matrimonio y afecta a otros muchos aspectos de sus relaciones.

De aquí que sea un tema de interés estudiar empíricamente quién toma las decisiones en la familia española actual, el marido o la mujer, respecto a una serie de puntos significativos de los aspectos de la vida familiar, y la relación del poder de decisión en estos puntos con las características personales principales de ambos cónyuges.

Indicar las hipótesis que se podrían adelantar en relación a esta cuestión a efectos de su comprobación empírica.

RESPUESTA

Puesto que entre los factores que pueden influir, en igualdad de circunstancias, en el poder de decisión en el matrimonio, se encuentran: la independencia econó-

mica de la mujer, la edad de los cónyuges, absoluta y relativa, el "status" ocupacional del marido, su nivel educativo, el ajuste marital y el tipo de personalidad del marido y de la mujer, etc., se podrían formular las siguientes hipótesis:

El poder del marido en la familia es mayor cuanto más elevado es su "status" ocupacional y nivel educativo, su personalidad es más autoritaria cuando el matrimonio es más reciente y su edad no es muy avanzada en relación a la de la mujer.

El poder de decisión de la mujer en el matrimonio tiende a ser más acusado cuando la mujer trabaja en una profesión y consiguientemente obtiene unos ingresos económicos, el marido tiene más edad que ella, llevan ya bastantes años casados y su satisfacción marital es deficiente.

El poder de decisión comparativo del marido y la mujer está en proporción con el "status" ocupacional de ambos, su edad, y la duración del matrimonio, su nivel educativo, su tipo de personalidad y el grado de ajuste sexual en el matrimonio.

13

Una pequeña comunidad rural mejicana ha rechazado el empleo de semillas de fríjoles mejoradas. Por ello se proyecta una investigación sociológica cuyo objeto sería hallar el motivo de este rechazo.

Enunciar una serie de hipótesis como explicaciones alternativas del fenómeno indicado y discutir cuáles de ellas deben ser elegidas como más probables para la investigación, teniendo en cuenta que se trata de una comunidad que ha admitido ya varios cambios técnicos como la luz y la radio de transistores, y que se trata de una medida patrocinada por el Gobierno, en relación al cual no existe una oposición generalizada, y llevada a cabo por sus técnicos.

RESPUESTA

Los motivos pueden derivarse de las características de la comunidad en cuestión, o de los organismos que han propuesto el cambio de semillas y lo han patrocinado, o de los agentes encargados de llevarlo a efecto.

Teniendo en cuenta lo anterior y los datos del enunciado se podrían formular las siguientes hipótesis alternativas:

1. La comunidad es supersticiosa y por eso ha rechazado las semillas mejoradas de fríjoles.

2. El cacique de la comunidad cree que perdería su poder y debido a su influencia la comunidad ha rechazado las semillas.

3. El técnico que fue enviado a la comunidad ignora la forma de tratar a su gente y por eso la comunidad ha rechazado las semillas en cuestión.

4. La comunidad es profundamente empírica y por eso ha rechazado las semillas que no ha experimentado personalmente y de la que no tiene ninguna experiencia.

5. El régimen político del país no satisface a la mayoría de la comunidad y por eso rechazan una ayuda que creen puede comprometer su independencia política.

6. La actuación y ayuda del Gobierno en otros campos no han sido muy eficaces y por eso no se tiene confianza tampoco en el resultado de este proyecto.

La observación del campo de investigación serviría para ir descartando algunas de estas hipótesis hasta quedarse con las más probables. Por ejemplo, el hecho de haber aceptado antes cambios técnicos como la luz eléctrica y la radio indica que son muy poco probables las hipótesis 1 y 4. Igualmente, la no existencia de una oposición generalizada al régimen político existente hace también poco probable la hipótesis 5.

(Basado en Felipe Pardinas: "Metodología y técnicas de investigación en ciencias sociales" Siglo XXI editores. Méjico, 1969.)

14

Dadas las siguientes hipótesis:

1. Si aumenta la prosperidad nacional, las clases medias se expanden.

2. Si las clases medias se expanden, aumenta el consenso de valores.

3. Si aumenta la movilidad social, el consenso de valores aumenta, y viceversa (Hans Zettemberg, 1965, pág. 83). Formular las nuevas hipótesis que se pueden deducir mediante la combinación de las anteriores teniendo en cuenta que la variable segunda de la primera hipótesis es la misma que la primera variable de la segunda hipótesis, que la variable segunda de las hipótesis segunda y tercera es la misma y que la hipótesis tercera es reversible.

RESPUESTA

Expresando matemáticamente la relación entre las dos variables de una hipótesis como una igualdad, se tendría en la primera hipótesis A = B, y en la segunda igualmente B = C. Pero, como es sabido, cuando dos miembros son iguales a un tercero son iguales entre sí; se puede concluir, por tanto, que A = C o lo que es igual que existe una relación entre las variables A y C que da lugar a una nueva hipótesis.

Esta sería:

Si aumenta la prosperidad nacional (A), aumenta el consenso de valores (C).

Aplicando el mismo razonamiento a las hipótesis 2 y 3 que tienen también un miembro común, el B, se tendría: A = B; C = B o B = C por ser reversible, luego A = C, igualdad que da lugar a la siguiente hipótesis:

Si las clases medias se expanden (A), aumenta la movilidad social (C).

15

Dadas las siguientes proposiciones o hipótesis:

1. Los grupos se renuevan menos que el público.

2. El público muestra menos emoción que la muchedumbre.

3. Los grupos muestran menos emoción que las masas, y teniendo en cuenta que:

a) Los grupos son agregados sociales que interactúan en términos de roles específicos con un líder común.

b) Las masas son agregados sociales que interactúan en términos de roles no especificados, pero con un líder común.

c) El público es un agregado social que interactúa en términos de roles específicos y sin un líder común, y

d) Las muchedumbres son agregados sociales que interactúan en términos de roles no especificados y sin un líder común.

Formular algunas nuevas hipótesis que se podrían deducir de las primeras proposiciones, utilizando en lugar de los términos las notas derivadas de sus definiciones.

RESPUESTA

En relación a la primera hipótesis o proposición, se ve que la diferencia entre grupos y públicos en la definición dada es que los primeros tienen un líder común y los segundos no, luego sustituyendo en la hipótesis propuesta el término grupo por esta noción diferencial, se tiene la formulación de la nueva hipótesis siguiente:

"Si un agregado tiene un líder común (grupo), su índice de renovación es menor que si el agregado no tiene un líder común (público)"

Igualmente en las otras dos hipótesis se ve que la diferencia en las definiciones dadas entre grupos y público, por una parte, y masas y muchedumbres, por otra, es que los primeros interactúan en términos de roles específicos, mientras que los segundos no. Por tanto, sustituyendo en las hipótesis propuestas 2 y 3, los términos grupos y públicos y masas y muchedumbres por su noción diferencial, se tendría la hipótesis siguiente:

"Si un agregado social interactúa en términos de roles específicos su nivel de emoción es menor que si interactúa en términos de roles no específicos."

(Basado en Haz Zettemberg, 1965, pág. 81.)

16

Formular cinco hipótesis en todas las cuales aparezca como variable dependiente o efecto los conflictos entre padres e hijos y como variable independiente los diversos factores que se estime pueden influir significativamente en la delincuencia juvenil.

RESPUESTA

1. A mayor rigidez ideológica de los padres, mayor proporción de conflictos entre padres e hijos.

2. A menor ajuste entre los padres, más conflictos entre ellos y los hijos.

3. Cuanto más liberal es el tipo de educación de los hijos, menores conflictos entre ellos y los padres.

4. A mayor intervencionismo de los padres en los asuntos de los hijos, mayor proporción de conflictos entre ellos.

5. A mayor diferencia de edad de los padres con los hijos, mayor proporción de conflictos entre ellos.

17

Formular cinco hipótesis en todas las cuales aparezca como variable independiente la presión social y como variable dependiente o resultado los diversos efectos que se estime puede producir la presión social.

RESPUESTA

Cuanto mayor es la presión social, mayor es la estabilidad familiar en una comunidad.

Cuanto mayor es la presión social, menor es la movilidad social.

Cuanto mayor es la presión social, mayores son las prácticas religiosas externas y ritualistas.

Cuanto mayor es la presión social, mayor es el número de desviados en una comunidad.

Cuanto mayor es la presión social, mayor es la proporción de enfermedades mentales en una comunidad.

18

Partiendo de que, como se suele admitir, la organización científica, típica de la sociedad moderna, se distingue por las siguientes características, que posee en alto grado: racionalización, especialización, coordinación, burocratización y mecanización, se pide formular hipótesis poniendo en relación estas características entre sí.

RESPUESTA

En realidad, este ejercicio es una aplicación del anterior 12, por cuanto el hecho de que se admita que la organización científica se distingue por las características indicadas, supone que todas ellas están en relación directa con aquélla.

Luego, por ejemplo, se puede admitir como hipótesis que si la racionalización y la especialización están relacionadas directamente con la organización científica, también lo están entre sí. Por tanto, se puede expresar esta relación en forma de hipótesis, y lo mismo con las restantes características, en forma similar a la siguiente:

A mayor racionalización, mayor coordinación.

A mayor racionalización, mayor especialización.

A mayor especialización, mayor burocratización, y viceversa.

A mayor mecanización, mayor burocratización, etc.

19

Parece que es un hecho, por lo general, la clara tendencia progresista de los sacerdotes católicos salidos del seminario en los años anteriores inmediatos al Concilio a pesar del tipo de formación tradicional entonces vigente en dichos centros.

Supuesto este hecho, se pide formular una serie de hipótesis alternativas que constituyan posibles explicaciones de este fenómeno desde un punto de vista social.

RESPUESTA

Los motivos explicativos de este hecho supuesto pueden tener su origen bien en las condiciones personales de los seminaristas, bien en sus condiciones familiares, bien en las circunstancias propias de los seminarios, bien en las influencias del ambiente social general de la época.

En consecuencia se podrían formular las siguientes hipótesis:

1. En los seminaristas de esta época predominaba un tipo de personalidad independiente y liberal y por eso adoptaron fácilmente una posición progresista.

2. Estos seminaristas procedían de familias muy tolerantes y abiertas a las nuevas corrientes y por su educación en ellas estaban predispuestos al progresismo.

3. En los seminarios predominaron en esta época líderes de tendencia autónoma y liberal, que impusieron la supremacía de esta corriente.

4. En los seminarios de esta época se dio un porcentaje relativamente elevado de vocaciones tardías, quienes con su formación secular y mundana influyeron en los demás.

5. En este tiempo penetró en los seminarios la literatura y las ideas de la dirección autocrítica del catolicismo y prendió en los seminaristas.

6. El progresismo en cuestión es una consecuencia de la penetración en los seminarios de las tendencias y orientaciones predominantes en el ambiente social de su época.

20

Uno de los fenómenos que es bastante probable se dé en los casos de rápido desarrollo económico es la intensificación de los procesos de la movilidad social, de tal forma que se puede formular la hipótesis siguiente:

"Cuanto más elevada es la tasa de crecimiento económico en una comunidad, tanto más aumentan en amplitud e intensidad los procesos de movilidad en ella"

Formular, partiendo de esta hipótesis, un conjunto de subhipótesis que la expliciten y concreten en relación a los diversos aspectos de la movilidad social, de modo similar a como se hizo en el ejercicio 17 número 6.

RESPUESTA

Habida cuenta de que el proceso de movilidad social, en los casos de rápido desarrollo económico, suele ser sobre todo importante, geográficamente, de las zonas rurales a las urbanas, y en éstas según se hallen más o menos industrializadas; en los individuos que tienen una mayor cualificación profesional, y, por último, en los jóvenes que en los viejos, algunas de las subhipótesis que se podrían formular son las siguientes:

1. Cuanto mayor es el ritmo de desarrollo económico de una comunidad nacional, tanto más intensa es la movilidad geográfica del campo a la ciudad.

2. En los procesos de desarrollo económico la movilidad vertical será tanto más amplia e intensa cuanto mayor sea el grado de industrialización de las ciudades o regiones.

3. En los procesos de desarrollo económico es probable que la movilidad vertical sea más intensa y amplia en los jóvenes que en los viejos.

4. Cuanto más elevado es el ritmo de desarrollo económico, tanto mayor es la amplitud e intensidad de la movilidad intergeneracional.

5. A mayor cualificación profesional, mayor grado de movilidad social vertical, en los procesos de desarrollo económico.

6. En los casos de rápido desarrollo es probable que la movilidad tenga lugar de modo más intenso en los sectores de la industria y los servicios.

21

Dando por admitido el siguiente principio o hipótesis abstracta de la teoría de Durkheim del suicidio egoísta, según el cual "un grado mayor de individualismo dentro de un grupo social origina una tasa mayor de suicidio en ese grupo", siendo el individualismo una variable que significa aproximadamente "el grado en el que todas las actividades de la persona son controladas por normas bien definidas, observadas regular y efectivamente por la gente en el medio ambiente, y en oposición a la moral determinada por el individuo mismo mediante sus propias decisiones", derivar, empleando la técnica de argumentación "si ... entonces", de dicho principio una serie de hipótesis de investigación o de trabajo.

RESPUESTA

Partiendo de dicho principio el mismo Durkheim utilizó las siguientes hipótesis de trabajo, que Stinchcombe llama "proposiciones empíricas", que son según él las que postulan que "si realizamos tales y tales observaciones, se obtendrán tales y tales resultados".

Si un grado mayor de individualismo dentro de un grupo social origina una tasa mayor de suicidios en ese grupo ....

a) ... entonces en Francia los protestantes tendrán tasas de suicidio más altas que los católicos;

b) ... entonces en las provincias alemanas, las regiones protestantes tendrán tasas más altas que las regiones católicas;

c) ... entonces en Francia, los hombres casados tendrán tasas de suicidio más bajas que los solteros, y éstas serán aun más bajas si tienen hijos:

d) ... entonces los hombres que practican profesiones liberales y en general los hombres cultos tendrán tasas de suicidio más altas que los obreros o que la gentes menos culta. (Basado en Stinchcombe, 1970, ps. 24-5).

6. UNIDADES DE OBSERVACION Y VARIABLES

NOCION Y CLASES DE UNIDADES DE OBSERVACION

De forma descriptiva, se puede afirmar que las hipótesis expresan mediante términos de unión relaciones entre variables referentes a unidades de observación determinadas. Esta noción enumera los elementos estructurales de las hipótesis desde el punto de vista científico, ya indicados al tratar de las hipótesis: las unidades de observación, las variables y las relaciones que las unen entre sí. La investigación social estudia, respecto a las unidades de observación, determinadas modalidades y características que conforman las variables.

Las unidades de observación son las realidades que se pretenden observar. Como tales constituyen en la investigación el objeto global de estudio y de ellas se obtienen los datos empíricos necesarios para contrastar las hipótesis con la realidad. Sin embargo, hay que tener en cuenta que, aun formando las unidades de observación las realidades globales estudiadas, en su realización práctica, la investigación científica y también la social no tienen lugar mediante el examen de los seres o cosas en su conjunto, sino a través del análisis de características determinadas de estas realidades, o sea de variables.

Como destaca Galtung, la elección de las unidades de observación es probablemente la primera elección decisiva que se hace en muchas investigaciones. Una vez realizada es difícil volverse atrás porque todo el proceso de investigación se habrá edificado sobre esta elección.

En cuanto objeto de observación y fuente de datos, las unidades de observación son entidades sustantivas que forman unidades de realidad más o menos independientes y que como tales poseen determinados atributos y propiedades y son susceptibles de diversos cambios y relaciones referentes a dichos atributos y propiedades.

No obstante lo anterior, es preciso subrayar, de modo similar a la aclaración antes expuesta, según la cual la investigación social se centra en las variables y no en las unidades de observación, que éstas no se confunden siempre con el objeto específico de la investigación. El objeto específico de investigación será frecuentemente un acontecimiento, por ejemplo, una huelga; un proceso: la movilidad social; un tipo de relación: la interacción social; una forma de cambio: la industrialización. En todos estos casos, dichas realidades no constituyen unidades de realidad independientes sino que son acontecimientos, procesos relaciones y cambios, que se producen respecto a determinadas unidades de observación: individuos, grupos sociales, etc.

Tampoco se confunden las unidades de observación con los hechos, considerados como toda realidad independiente del pensamiento, en cuanto que son hechos también las propiedades, cambios, procesos y relaciones, es decir, las variables.

Aclarado lo anterior, las unidades de observación fuentes principales de datos en la investigación social son:

- Los individuos.
- Los grupos sociales.
- Las instituciones, y
- Los objetos culturales, obra del hombre.

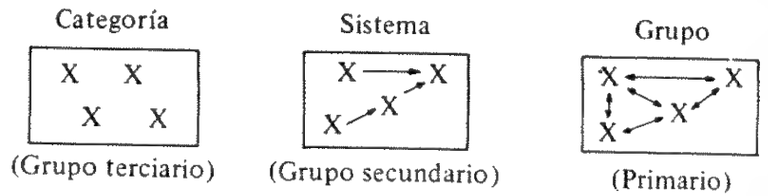
Aunque los individuos humanos aisladamente no sean objeto global de estudio directo e inmediato en las ciencias sociales, sí son frecuentemente unidades de observación en estas ciencias. Es obvio que en muchos estudios sociológicos la unidad de observación es el individuo. En ellos se obtiene información sobre las sociedades, observando a poblaciones o muestras de individuos y reuniendo los resultados individuales. Es decir, se procede mediante la agregación de datos individuales.

Con relación a este punto, hay que distinguir también entre unidades de observación y unidades de análisis. En las investigaciones sociales, las unidades de observación pueden ser y de hecho lo son a menudo individuos; pero las unidades de análisis no son individuos, sino agregados de individuos, formados, por ejemplo, por los que votan al partido X, los contrarios a la pena de muerte, etc.

Las unidades de observación en la investigación social se dividen en individuales y colectivas y éstas, a su vez, se pueden subdividir con Galtung, en categorías, sistemas y grupos.

La categoría es un conjunto de unidades sin estructura, es decir, sin interrelación entre sí. En el sistema todas las unidades están conectadas pero no necesariamente de forma directa.

El grupo es un sistema fuertemente interrelacionado. En él todas las unidades están directamente conectadas entre sí.



La selección de las unidades de observación se debe basar en una idea clara:

1. De las propiedades, relaciones o procesos objeto de investigación.
2. De a qué entidades se refieren éstas y por tanto deben constituir las unidades de observación.
3. De las características de las unidades de observación y de su grado de estructuración cuando sean colectivas.

NOCIONES E IMPORTANCIA DE LAS VARIABLES

Como se ha indicado en el parágrafo anterior, las hipótesis expresan, mediante términos lógicos, relaciones entre variables referentes a unidades de observación determinadas. De estos elementos que forman las hipótesis, las variables son particularmente importantes en la investigación científica.

Las variables presentan, tomadas separadamente, dos características fundamentales: primero, ser características observables de algo, y segundo, ser susceptibles de cambio o variación con relación al mismo o diferentes objetos. Pero si se consideran no aisladamente, ofrecen además la particularidad de presentarse a la observación vinculadas en su variación unas a otras, según muy diversos tipos de relaciones.

De aquí que se pueda distinguir una doble noción de variable. Según la primera, las variables son, como las define Linton C. Freeman (1971, página 18), "características observables de algo que son susceptibles de adoptar distintos valores o de ser expresadas en varias categorías". Por ejemplo, una variable es el color que pueden presentar las categorías de blanco, negro, verde, azul, etc., y otra el peso de una persona que puede tomar los valores de 60 kilogramos, 70, 80, etc.

Conforme a la segunda noción, las variables son características observables de algo, ligadas entre sí en su variación con una relación determi-

nada, entre las que se pueden citar como las más comunes las de covariación o asociación y de dependencia, influencia o causalidad. Por ejemplo, la variable posición social de una persona está ligada en una relación de dependencia con diversas variables, entre las que se pueden señalar su edad, sexo, ingresos, poder, inteligencia y nivel educativo.

La primera noción tiene validez especialmente en los estudios puramente descriptivos en cuanto es su finalidad determinar o medir las categorías variables que caracterizan a un fenómeno. En cambio la segunda tiene aplicación en los estudios que, sean o no propiamente explicativos, pretenden estudiar las relaciones, los motivos, razones y efectos de los fenómenos sociales.

La importancia de las variables en el método científico y, por tanto, también en el sociológico es básica. La investigación científica gira alrededor de ellas. La finalidad del trabajo científico no es otra que descubrir la existencia de las variables y su magnitud y probar las relaciones que las unen entre sí. Pero no sólo constituyen el fin inmediato y directo de la investigación científica según lo anterior, sino que forman también su objeto, en cuanto se opera con ellas en todas las fases de la investigación, en la observación, clasificación, análisis y explicación, de las que constituyen los elementos clave y los términos operativos.

En efecto, se opera con ellas en todas las fases principales de la investigación. En la observación se las descubre y mide. En la clasificación se las agrupa. En la tabulación se las relaciona. En el análisis se las estudia e interpreta.

Hay que tener en cuenta que la noción de variable presente, en principio, la máxima generalidad. Esto quiere decir que incluso todas las cosas y realidades del mundo, que constituyen el objeto de la ciencia, y no sólo sus características, pueden ser consideradas como variables porque pueden ser categorías de realidades más amplias que las engloban. Por ejemplo, España además de ser una unidad de observación es también una categoría o elemento de variación de la variable naciones europeas. Por otra parte la posibilidad de su existencia o no existencia, da lugar a que toda realidad pueda ser en este sentido una variable con dos categorías: la afirmación de la cosa y su negación. Por ejemplo, obrero y no obrero.

En este sentido la atribución del carácter de unidad de observación o de variable es frecuentemente algo relativo y circunstancial, que depende de cada caso y situación, y del nivel de generalidad en que nos situemos.

J.A. Davis (1971, p. 9) distingue en las variables los siguientes cuatro elementos:

a) un nombre,
b) alguna especie de definición verbal, o, de modo más preciso, operativa.
c) un conjunto de categorías,
d) un procedimiento para realizar la clasificación de las unidades de observación en las categorías.

Ejemplo: Nombre de la variable: Preferencias de partido. Definición verbal: Voto usual en las elecciones nacionales. Categorías: Demócrata, republicano, otro, ninguno. Procedimiento de clasificación: Respuesta de los encuestados a esta pregunta. ¿En las elecciones nacionales a favor de qué partido vota usualmente? (Ver ejercicio 27)

LAS UNIDADES DE OBSERVACION Y LAS VARIABLES COMO CONCEPTOS Y TERMINOS

En las unidades de observación es preciso distinguir, por una parte, las unidades de observación como entidades reales y las variables como sus características y, por otra, los conceptos mediante los cuales aprehendemos intelectualmente dichas realidades y los términos lingüísticos que utilizamos para expresarlos. Es obvio que en las operaciones de investigación se trabaja con los conceptos y términos y no con las realidades de las unidades de observación excepto en ésta.

Según se expuso al tratar de la estructura de las hipótesis, las unidades de observación y las variables, como entidades reales y sus características, corresponden al nivel científico o empírico, pero, desde el punto de vista lógico, no son otra cosa que conceptos representados lingüísticamente por términos verbales.

LOS CONCEPTOS Y TERMINOS Y SUS NIVELES

Concepto significa literal y etimológicamente, lo concebido, término que, dada la similitud del conocimiento con la generación y el nacimiento, expresa adecuadamente que el concepto es el nuevo ser intelectual resultado inmediato y primero de la actividad cognoscitiva.

Por esta circunstancia de ser el resultado inmediato y primero del conocimiento, son también los conceptos su forma más simple y por ello, como hemos indicado, los materiales básicos de todo tipo de conocimiento, incluido, por tanto, el científico. Por ser la forma más simple del pensamiento, se limita a significar una realidad, sin afirmar o negar nada de ella, por ejemplo, el concepto libro representa el significado

de libro pero no nos dice nada más de él ni siquiera implica necesariamente su existencia.

El concepto en sí, su contenido, es la idea o el significado que representa, pero necesita también, como todo tipo de conocimiento, de una forma generalmente lingüística, mediante la cual adquiere expresión que, en el caso del lenguaje, recibe el nombre de término.

Los elementos fundamentales, pues del concepto con su contenido o significado y el término que da forma lingüística al concepto.

A pesar de esta unión entre el concepto como significado y el término lingüístico, ambos no coinciden necesariamente en cada caso concreto, pues frecuentemente se tienen para expresar el mismo concepto varios términos diferentes y viceversa.

El significado le viene al concepto de su referencia a una realidad distinta e independiente del concepto, que es su referente. Por ello en relación a los conceptos se pueden distinguir, como lo hace Bunge (1972, 76.7) tres niveles:

- El nivel real de su referente, p. ej. una sociedad determinada.
- El nivel propiamente conceptual de la idea o significado que se adopta de la sociedad y que puede representar más o menos exactamente a la sociedad real, y, por tanto, ser más o menos objetivo, y, en fin,
- El nivel lingüístico, que es el término o la palabra sociedad con la que designamos el concepto de sociedad.

La relación que existe entre el nivel lingüístico y el conceptual es de designación: el término designa un concepto, y entre el nivel conceptual y el real, de referencia; el concepto se refiere a una realidad.

LA INTENSION Y LA EXTENSION

En los términos se pueden distinguir dos elementos fundamentales: su intensión y su extensión. La intensión significa el contenido del término, el conjunto de propiedades y relaciones que comprende.

La extensión es el conjunto de sujetos a los que se aplica o puede aplicar el término.

La mayor o menor intensión y extensión de los términos determina su grado de abstracción o concreción, es decir, de separación o cercanía respecto a la realidad.

Los términos se pueden diferenciar, por tanto, por su grado de abstracción. Sociológicamente cuanto menos abstracto es un concepto más cercano se halla de los procedimientos de observación y de medida actuales.

Sean por ejemplo, los términos: clase social, clase alta y directores de grandes empresas. El concepto de clase social que es el más abstracto de los tres, es también el más amplio en extensión y el menos en intensión.

El significado en extensión e intensión de un cierto término no figura de manera expresa en el término o signo sino que es preciso descubrirlo mediante su análisis.

En relación a la concreción y a la abstracción, se da un movimiento inverso en la intensión y extensión de los términos.

Respecto a la intensión, los términos conceptuales, cuanto más concretos son, más propiedades o determinaciones comprenden, pues abarcan no sólo las del género y la especie a que pertenecen, sino también, las peculiaridades de un individuo determinado.

En cambio, la extensión es mínima en los términos concretos, individuales, y máxima en los términos abstractos, por ejemplo la humanidad.

Al constituir los términos, las unidades de observación y las variables de las proposiciones científicas y de las hipótesis tienen una importancia básica en la investigación social.

Como dice el profesor Salustiano del Campo en el Diccionario de Ciencias Sociales del IEP, "los conceptos son elementos tan básicos del método científico como las hipótesis; es más, éstas no podrán existir sin aquéllos. El sistema teórico de la ciencia es un sistema conceptual".

DEFINICION DE CONCEPTOS

Según Suppues (1957, p. 152) una definición es un enunciado que establece el significado de una expresión. Se realiza refiriendo el término definido, el *definiendum*, a otros términos, los *definiens*, ya disponibles.

La definición previa de los términos conceptuales que forman las unidades de observación y las variables es una necesidad básica en toda investigación científica. Mediante ella, se elige la noción que se estime más adecuada a los fines pretendidos, y se precisa el sentido único en que se han de tomar las unidades y las variables por todos los que intervengan en la investigación, a la vez que se hace factible la posibilidad de la exacta comprensión del sentido y alcance de la investigación a otros especialistas y al público en general.

La definición de las unidades de observación debe realizarse de tal forma que sean perfectamente identificables en la realidad tales unidades y se evite la confusión con otras similares, para no caer en errores de atribución.

En cuanto a la definición de los términos de las variables se requiere además, que se basen en una referencia a los rasgos que presentan en la realidad, para que sean susceptibles de operativización, como se verá más adelante.

Se pueden citar cuatro tipos principales de definiciones: reales, nominales, ostensivas y operacionales.

Las primeras son las que pretenden exponer la naturaleza o esencia de una cosa, mediante la formulación de su género próximo y su diferencia específica. Por ejemplo: la definición de hombre como animal racional.

Las definiciones nominales prescinden de la pretensión difícil de lograr indicar la naturaleza, lo que sea una cosa en sí, y se limitan a establecer lo que se entiende por, o el sentido en que se va a tomar en una investigación, trabajo o exposición, el término o palabra en cuestión.

Definición ostensiva es la que se realiza mostrando o indicando un individuo de la clase de los objetos a que se refiere una definición. Por ejemplo, mesa es esta mesa que tenemos delante.

Especial importancia tienen en la investigación social las llamadas definiciones operativas.

Con relación a las ciencias sociales, Blalock (1966, p. 21) afirma que "las definiciones operativas son definiciones que enuncian efectivamente los procedimientos empleados en la medición –y por tanto, se puede añadir, también en la observación, dado que, como veremos, esta última se puede considerar como el resultado de la medición—. La definición operativa "longitud", p. ej., "indicará exactamente como deba medirse el largo del cuerpo".

Así pues, mientras las definiciones reales y nominales se pueden considerar como definiciones teóricas y abstractas, dado que en ellas unos términos o conceptos se definen por otros que se dan por comprendidos, las definiciones ostensivas y operativas serían definiciones empíricas y concretas, en cuanto basadas en la observación y medida referidas a indicadores de la realidad directamente observables.

A pesar de ser la definición de términos un punto importante, sin duda, se halla muy poco estudiado y elaborado en la investigación social.

Ch. Lachenmeyer sostiene (1976, p. 41 y ss) que los términos que se usan en Sociología adolecen frecuentemente de vaguedad (no se sabe bien cuáles son sus límites, qué es lo que abarcan), de ambigüedad, porque se emplean con muy diversas significaciones y de opacidad porque no está claro cuál es su referente empírico, la realidad a la que se refieren. (Ver ejercicios 22, 27)

Una forma práctica de proceder al respecto podría ser la siguiente que expone Ackoff (1973, p. 55 y ss):

1. Examinar tantas definiciones del término pasadas y presentes como sea posible. Conservar la cronología de las definiciones.
2. Intentar penetrar en el núcleo de significación hacia el que la mayoría de las definiciones parecen apuntar.
3. Formular una definición tentativa basada en dicho "núcleo".
4. Ver si este intento de definición cubre todos los casos que se piensa debería cubrir, respecto a los objetivos de la investigación.
5. Someter esta definición a una valoración tan crítica como sea posible por parte de científicos y no científicos.
6. Realizar una revisión final de la definición sobre la base de las críticas legítimas que se reciban.

FORMACION DE CONCEPTOS

En la formación de los conceptos hay que distinguir como hasta aquí la formación de los conceptos mismos y la de los términos que los representan.

La formación de los conceptos, en cuanto a su contenido ideal, es el resultado de la misma actividad cognoscitiva, incluida la sensación y la percepción, que no corresponde examinar aquí. En cuanto a la formación de los términos que representan los conceptos, es interesante para explicarla, la distinción que hace De Bono de los paquetes de ideas y las ideas con nombre propio.

Cuando se trata de comprender una situación extraña se pueden recapitular todos los aspectos y propiedades que percibimos en un paquete de ideas. Esta lista expresará aspectos del contenido del concepto, de su intensión y también de su extensión.

La utilización de este paquete de ideas del concepto de una cosa en la comunicación lingüística sería muy poco práctica y viable. De ahí la necesidad de atribuirle un nombre propio, un término que represente no

sólo el paquete de ideas sino incluso los aspectos y propiedades de la cosa no comprendidos en él.

La formación de conceptos en este sentido no consiste, pues, en otra cosa sino hallar un término, la palabra o palabras, que sustituya sintéticamente al paquete de ideas y represente la realidad conocida.

Desde el punto de vista lingüístico, mediante esta formación se ejercita la función denominadora del lenguaje por la que se atribuyen signos lingüísticos que representan las cosas, las cualidades, las actividades, las situaciones y las relaciones.

La formación de conceptos se concreta, según lo anterior, en la enumeración de las propiedades de la realidad innominada y en la búsqueda, en el acervo de la lengua empleada, de un término, de acuerdo con las características de ésta.

TIPOS DE VARIABLES

Las variables se pueden clasificar, entre otros modos según: a) su naturaleza; b) la amplitud de las unidades de observación a que se refieren, c) su nivel de abstracción, d) el carácter de los elementos de variación que comprenden y e) su posición en la relación que une a dos o más variables entre sí.

a) Según su naturaleza se pueden distinguir las variables cualitativas y cuantitativas. Las primeras son aquellas cuyos elementos de variación tienen carácter cualitativo o no numérico, y al contrario de las segundas en las que dichos elementos tienen carácter numérico o cuantitativo.

Las cuantitativas a su vez se pueden subdividir en agrupadas o no y en discretas y continuas.

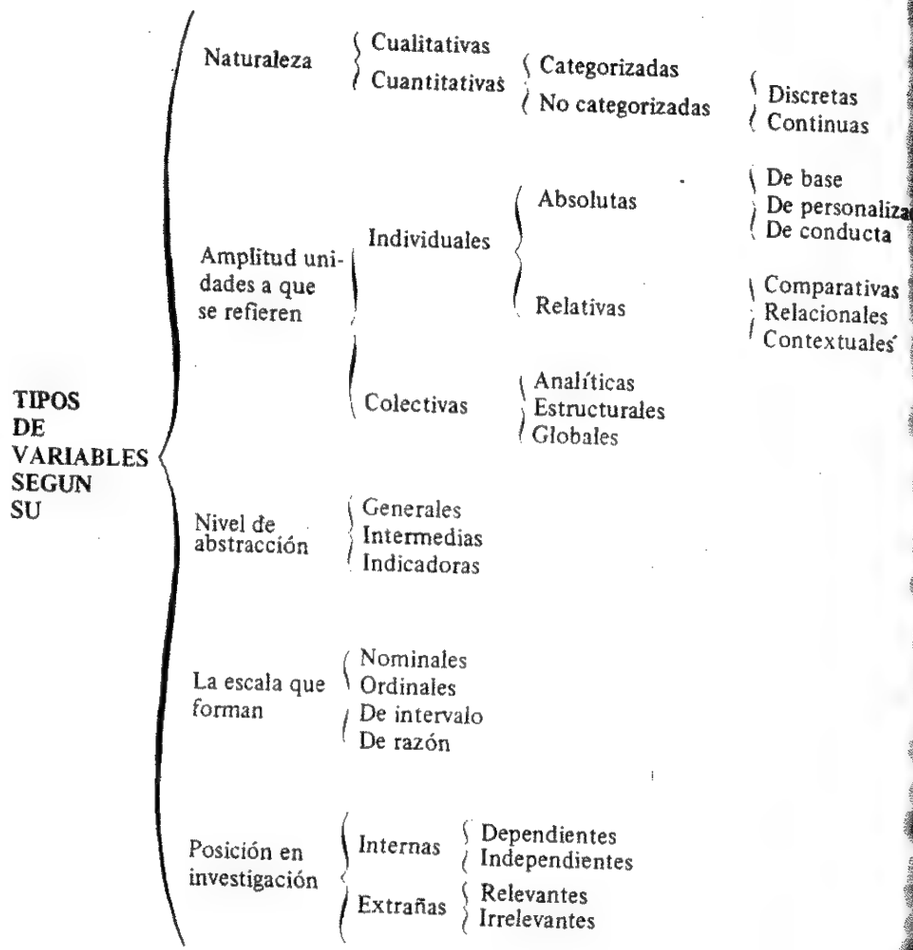
Las agrupadas son aquellas en que la serie numérica que abarcan está dividida, a efectos operativos y de síntesis, en intervalos agrupados.

Las continuas pueden tomar cualquier valor dentro de su rango, mientras que las discretas, por el contrario, se hallan restringidas a determinados valores dentro de su rango.

b) Según la amplitud de las unidades de observación se puede hablar de variables individuales y de variables colectivas.

Las primeras se refieren a características de las unidades de observación cuanto éstas son individuos.

Variables colectivas son las que se refieren a características de las unidades de observación cuando éstas son colectivos, conjuntos o grupos



(ciudades, estados, universidades, escuelas...) Ejemplo: Urbanización, tasa de fertilidad, industrialización. (Ver Ej. 27).

Una clasificación de las variables individuales y colectivas que merece mención especial por su interés analítico es la debida a Lazarsfeld y Menzel. Estos autores en un artículo de gran resonancia publicado en el año 1961, en A Etzioni (ed). *"Complex organisations"*, N. York, Holt, 1961) dividen las variables individuales en absolutas y relativas y éstas últimas a su vez en comparativas, relacionales y contextuales. Por su

parte en las colectivas, distinguen las variables individuales, estructurales y globales.

Las absolutas no implican referencia alguna a otra característica del individuo o grupo. Por ejemplo, las variables físicas: edad, peso, estatura; las conductas individuales y los rasgos psicológicos.

Relativas, son las que implican tal relación. Entre ellas, las comparativas tienen su origen en una comparación. Por ejemplo, ser el más viejo del grupo, poseer una inteligencia media en un grupo dado.

Las relacionales caracterizan al individuo por el tipo de su participación en una relación social. Son tales las variables que indican el rol de los individuos y la posición que ocupan en el grupo.

Las contextuales indican una relación entre el individuo y el medio ambiente social al que pertenece. Constituyen propiedades de un colectivo que se usan para caracterizar a sus individuos: P. ej.: recibir educación en un colegio privado.

Relacionadas con las variables contextuales están las variables ecológicas, que se refieren al medio ambiente físico en que se desenvuelven los individuos y grupos humanos.

Según el sistema de medida de las variables colectivas, se distinguen las variables analíticas cuando tal sistema se funda en los individuos, p. ej. el nivel económico de una sociedad medida por la renta de sus individuos, las variables colectivas globales en las que tal sistema de medida se basa en colectivos, p. ej. el mismo nivel económico medido por la tasa de establecimientos bancarios, y las variables estructurales, que se derivan de operaciones con datos de los miembros de un colectivo, referentes a sus relaciones sociales y su interacción. Por ejemplo, la cohesión de un grupo se mide en sociometría por el número de elecciones recíprocas dentro del grupo.

Dentro de las variables individuales absolutas se pueden distinguir con Boudon (1973, v. 1, p. 53), las variables de base, de personalidad y de opiniones y conducta.

La primera sitúa a los individuos en la estructura social y configura su personalidad social. Tienen carácter público y relativamente permanente. Ejemplos de ellas son la edad, el sexo, la profesión, el estado civil, la religión, etc.

Las variables de personalidad definen el temperamento, el carácter, el talante y otras cualidades personales del individuo. Son permanentes y privadas. A veces ni siquiera es consciente de ellas la persona.

El resto de las variables individuales absolutas caen dentro del tercer grupo. Son privadas y transitorias. Muchas de las preguntas de los cuestionarios son de este tipo. (Ver Ej. 27).

c) Por su nivel de abstracción, se tienen las variables generales que se refieren a realidades no inmediatamente medibles empíricamente, las variables intermedias que expresan dimensiones o aspectos parciales de estas variables y, por tanto, más concretos y cercanos a la realidad, y las variables empíricas o indicadores que representan aspectos de estas dimensiones directamente medibles y observables. El interés práctico de esta clasificación reside en que, como se verá, el proceso de hacer operativas las variables no consiste en otra cosa que en traducirlas en variables intermedias o indicadores. (Ver ejercicios 25, 26, 27, 28 y 29)

d) Según el carácter de las escalas o conjuntos que forman los elementos que comprenden las variables, éstas se pueden dividir en nominales, ordinales, de intervalo y de razón. Las variables de escala nominal son las que comprenden la distinción de diversas categorías como elementos sin implicar ningún orden entre ellas. Las variables de escala ordinal implican orden entre sus categorías, pero no grados de distancia iguales entre ellas. Las de intervalo suponen a la vez orden y grados de distancia iguales entre las diversas categorías, pero no tienen origen natural, sino convencional. Por último, las de razón comprenden a la vez todos estos aspectos, distinción, orden, distancia y origen único natural. Esta clasificación es sobre todo importante a efectos de análisis estadístico de los resultados obtenidos en la investigación. (Ver ejercicio 24 y 27)

e) Por último, conforme a su posición en la relación que une las variables entre sí, se dividen en dependientes e independientes. Las primeras designan las variables a explicar, los efectos o resultados respecto a los cuales hay que buscar su motivo o razón de ser. Las independientes son las variables explicativas, cuya asociación o influencia en la variable dependiente se pretende descubrir en la investigación. Esta distinción hace referencia al núcleo central del proceso investigador y, por tanto, es básica en él.

En muchos casos, dada la interacción entre ellas, no se puede determinar de modo absoluto en las variables sociales las que lo son dependientes o independientes. Frecuentemente que tengan un carácter u otro en una investigación, es convencional y depende del objetivo perseguido en la investigación.

Esta división de las variables en independientes y dependientes se refiere, obviamente, no a las variables aisladas sino a los enunciados o hipótesis en los que se relacionan variables entre sí de diverso modo.

Además de las variables dependientes e independientes, existen otras variables extrínsecas a la investigación o no investigadas. Si sus efectos no son importantes reciben el nombre de irrelevantes y si lo son de relevantes. (Ver Ejercicios 23 y 30).

TIPOS DE RELACIONES ENTRE VARIABLES DEPENDIENTES E INDEPENDIENTES

Estas relaciones pueden presentar diversas modalidades. En primer lugar merecen destacarse (Zettemberg, 1971) los siguientes tipos, que se enuncian en cada caso con la significación de la relación entre la variable independiente X y la dependiente, Y.

- *Reversible:* Si X, entonces Y, pero también si Y, entonces X.
- *Irreversible:* Si X, entonces Y, pero no viceversa.
- *Determinista:* Si X, entonces siempre Y.
- *Estocástica:* Si X, entonces probablemente Y.
- *De secuencia:* Si X, entonces más tarde Y.
- *Coexistente:* Si X, entonces también Y.
- *Suficiente:* Si X, entonces Y independiente de cualquier otra cosa.
- *Contingente:* Si X, entonces Y, pero sólo si Z.
- *Necesaria:* Si X, y sólo si X, entonces Y:
- *Sustituíble:* Si X, entonces Y, pero si Z, también Y.

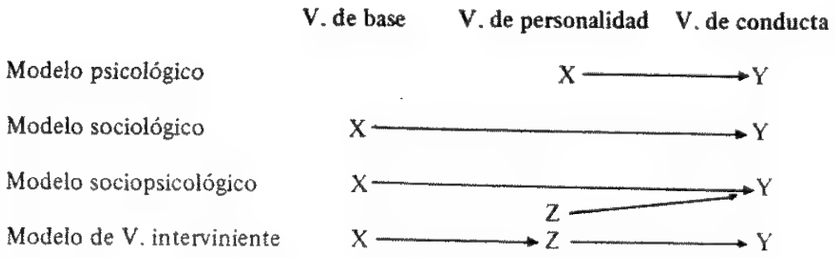
Las relaciones entre variables en las ciencias sociales son por lo general reversibles, estocásticas, contingentes y sustituíbles. Todo ello es signo de su debilidad y de la dificultad del estudio científico en las ciencias sociales de las relaciones entre variables y explica el bajo nivel logrado hasta ahora por dichas ciencias. (Ver ejercicio 23)

Otro aspecto de la relación entre variables, dependientes e independientes, es el tipo de modelo que se produce cuando se tiene en cuenta el tipo de variable dependiente e independiente. Galtung ha distinguido los siguientes modelos:

- *Psicológico:* En él, la variable independiente, es de personalidad y la dependiente, de conducta o de opinión.
- *Sociológico:* En él la variable independiente es de base y la dependiente de conducta o de opinión como antes.
- *Modelo socio-psicológico:* En él la variable de conducta o de opinión es influida a la vez por una variable de base y por otra de personalidad.

Otra modalidad de este modelo, sería el que Galtung llama de variable interviniente, en el que la variable de base influye en la de personalidad y ésta por último en la de conducta.

Además de estos modelos también se podría hablar de los modelos estructural, contextual y ecológico, cuando la variable base independiente es de uno de estos tipos y la dependiente de opinión o de conducta.



OPERATIVIZACION DE LAS VARIABLES

Al tratar de los distintos tipos de variables vimos que, según su grado de abstracción, se podían dividir en variables generales, intermedias y empíricas o indicadores. La operativización de variables, como se apuntó en dicho lugar, no es otra cosa que el procedimiento que tiende a pasar de las variables generales a las intermedias, y de éstas a los indicadores, con el objeto de transformar las variables primeras de generales en directamente observables e inmediatamente operativas. Esta transformación tiene un gran interés práctico, pues es la condición para que se puedan estudiar las variables generales, referentes a hechos no medibles directamente, y llevar a cabo la investigación propuesta sobre ellas.

El procedimiento al efecto ha sido desarrollado por Paul F. Lazarsfed (Cfr. Boudon, R. y otros Vol. 1) quien distingue en él las siguientes fases:

1.ª Representación del concepto de la variable, de modo que resulten expresados en una noción teórica los rasgos principales que presenta en la realidad.

2.ª Especificación del concepto hallado, o sea, el análisis de los aspectos y dimensiones de interés práctico, implicados en la representación del concepto de la variable, o derivados empíricamente de la estructura de sus intercorrelaciones. Por ejemplo, Lazarsfeld señala que en la noción de rendimiento, se pueden distinguir como componentes o dimensiones los tres siguientes: ritmo de trabajo, calidad del producto y rentabilidad del equipo.

3.ª Elección de indicadores, o de circunstancias empíricas concretas, que sean signo de la mayor o menor extensión que alcanza la dimensión en cuestión en las unidades de observación investigadas. Esto exige llevar la labor de concreción que la operativización supone hasta su límite máximo posible, buscando todos los indicios que en la realidad se puedan considerar como muestra de la dimensión de que se trata. Por ejemplo, una variable intermedia o dimensión de la variable general clase social, es el nivel económico, e indicadores de éste son el importe de todas las fuentes de renta: sueldos, rendimientos de fincas rústicas y urbanas, intereses de capitales, dividendos de acciones, ausencia de deudas, capacidad para obtener créditos, etc.

4.ª La cuarta fase es la construcción de índices. Una vez seleccionados los indicadores imaginados que parezcan más importantes y aptos para la investigación, es normal que no se pueda asignar a todos la misma importancia. De aquí la necesidad de construir un índice que agrupe en una medida común todos los indicadores referentes a una dimensión, mediante la asignación de un peso o valor a cada uno de ellos de acuerdo con su importancia.

Por tanto, en síntesis, para operacionalizar una variable, 1° se enuncia o define ésta; 2° se deducen sus dimensiones o aspectos principales; 3° se buscan indicadores de cada dimensión, y 4° se forma, en su caso, un índice.

Es preciso no olvidar que la operativización de variables consiste, en último término, en la sustitución de unas variables por otras más concretas. Por tanto, para que sea válida esta operación es necesario que las variables, que sustituyen a otra u otras, sean representativas de ella o de ellas.

La operativización de variables implica la adopción de una definición operativa de las variables investigadas. El primer requisito , como señala Paul C. Stern (1979), para que las observaciones tengan valor científico, es la concreción de las abstracciones, la cual se realiza adoptando una definición operativa para cada abstracción. Es necesario concretar, operativizar o definir operativamente las abstracciones para que se sepa exactamente qué se ha de observar. (Ver ejercicios 28, 29, 30, 72a, 76 y 87 a 89)

Se pueden preguntar por qué no prescindir de las variables intermedias y buscar directamente los indicadores, sin pasar por aquéllas previamente.

La contestación es que es la manera de proceder con orden y de abarcar todos los aspectos de la variable general, y además facilita la búsqueda de indicadores al concretar y subdividir su objeto.

Adviértase la importancia que tiene este procedimiento en las ciencias sociales. En ellas casi todas las variables investigadas son hechos frecuentemente internos, no aparentes, expresados por términos sin una referencia empírica externa.

Como decía Durkheim, para estudiar la solidaridad, un hecho interno, hay que sustituirla por indicadores que la simbolicen.

Esto es lo que hace la operacionalización de variables de una manera sistemática. Mediante ella, vamos construyendo una red que nos permita captar la realidad.

De la perfección de esta red, de que verdaderamente contenga todos los aspectos significativos de la realidad, depende el éxito de la investigación.

LAS VARIABLES EMPIRICAS O INDICADORES

Por su importancia práctica en la investigación social interesa considerar especialmente las variables empíricas o indicadores. Los indicadores reciben este nombre porque indican o son indicios de otras variables más generales y por ello de su existencia se puede inferir la concurrencia de dichas variables más abstractas de las que son signo y con las que están relacionados. En este sentido, Nowak (1977) las define como todo rasgo, suceso, fenómeno de cuya ocurrencia concluimos, con certeza o más o menos probabilidad, que el fenómeno que nos interesa tiene lugar.

Teniendo en cuenta a dicho autor y a Mayntz (1975) se pueden distinguir los siguientes tipos de indicadores:

- *Definicionales.* En ellos el indicador se identifica con lo que se quiere estudiar por medio de él. Son tales, por ejemplo, los indicadores sociométricos del número de elecciones de un individuo en un grupo y el % de elecciones recíprocas que definen los conceptos de status sociométricos y cohesión grupal.
- *Empíricos.* En este caso, el indicador y lo indicado están unidos por una relación empírica. Se subdividen en observacionales, si lo

indicado y el indicador son observables, como ocurre en el caso del humo, indicador del fuego; e inferenciales, si lo indicado es no observable o latente: ponerse rojo (observable) de vergüenza (no observable.

Los indicadores pueden constituir también un elemento entre un síndrome o conjunto de elementos observables de un fenómeno o, incluso, pueden no ser elementos en sentido estricto de dicho síndrome: por ejemplo, lectura de literatura de un cierto tipo, como indicador del gusto, por el arte abstracto.

Maintz llama indicadores correlativos, internos a los primeros y externos a los segundos. Los primeros, como partes de lo indicado, han de estar correlacionados con los restantes elementos del síndrome o conjunto de síntomas; los segundos también, aunque no formen parte de él.

Respecto a la relación de los indicadores con lo indicado, son posibles cuatro situaciones:

1. Que el indicador e indicado coincidan;
2. Que el indicador esté contenido en lo indicado;
3. Que el indicador contenga a lo indicado;
4. Que se intersecten.

En relación a estas situaciones, Nowak define las medidas siguientes:

- *Poder de rechazo, PR.* Indica la probabilidad de que un objeto que no pertenece al indicado C sea indicado por el indicador I.

$$PR = IC/I$$

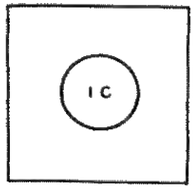
- *Poder de inclusión, PI.* Indica la probabilidad de que un objeto que pertenece al indicado C, esté indicado por el indicador I.

$$PI = IC/C$$

- *Poder discriminante, PD.* Es la correlación entre indicador e indicado.

$$PD = r_{IC,}$$

I es en estas fórmulas, la extensión del indicador I, C, la de lo indicado e IC la coincidencia de la extensión de I con la de C.

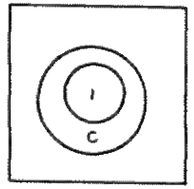


Situación I

PR = 1

PI = 1

PD = + 1

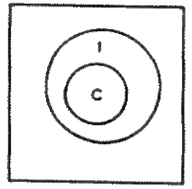


Situación 2

PR = 1

PI < 1

0 < PD < + 1

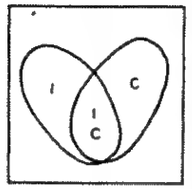


Situación 3

PR < 1

PI = 1

0 < PD < + 1



Situación 4

PR < 1

PI < 1

– 1 < PD < + 1

El indicado es el fenómeno que nos interesa conocer, y que, por tanto, constituye el objeto principal de nuestra investigación, y el indicador es el que realmente observamos. Cuanto más débil sea la correlación entre ambos, más grande será el riesgo de que los resultados de la investigación sean infundados respecto al indicado.

En cuanto al poder de inclusión o rechazo, a veces podemos estar interesados en indicadores con alto poder de exclusión, otras de inclusión y otras de ambos.

A veces es conveniente utilizar dos indicadores: uno, con un alto poder de rechazo y otro con alto poder de inclusión.

EJERCICIOS

22

En los Estados Unidos se realizó una investigación para comprobar empíricamente si existía discriminación en los restaurantes de Nueva York respecto a los clientes negros. Se partió de la hipótesis de la posible existencia de algún tipo de descriminación hacia los clientes negros en dichos restaurantes.

Efectuar la definición de términos de la unidad de observación y de las variables.

RESPUESTA

La unidad de observación son los clientes negros. La investigación no puede basarse en una definición técnica de clientes negros, sino en una que sea aplicable en la práctica y utilizable para distinguir, mediante la observación, los clientes negros de los demás. Podría ser ésta: "toda aquella persona que, por el color de su piel y sus rasgos físicos, pueda estimarse que pertenece a la raza negra".

La variable sólo es una, la discriminación en los restaurantes de Nueva York de los clientes negros. El término a definir, por tanto, es qué se debe entender por discriminación en restaurantes.

Discriminación en general, en este caso como en todos, es desigualdad de trato. Esta noción general es preciso especificarla de acuerdo con el objeto de la investigación, según el cual ha de referirse concretamente a clientes o comensales negros en los restaurantes, y por motivos precisamente raciales, y no de otro tipo.

Por ello se definió la discriminación en el trabajo en cuestión como "cualquier desigualdad de trato concedido a comensales negros y blancos, a menos que haya razón para creer que la diferencia en el trato es debida a factores distintos de la diferencia de raza". (Vid. C. Selltizt, 1970, pág. 88.)

Esta definición no es suficientemente concreta porque la desigualdad de trato puede adoptar muchas modalidades. Hay, pues, que especificar qué desigualdades concretas se tomaron en consideración, tales como situarles en sitios aparte y menos lujosos que los blancos; servir primero a los blancos que a los negros, aunque éstos hayan llegado antes, etc.

23

En los enunciados hipotéticos siguientes:

I) En una sociedad dada, el absentismo laboral es más frecuente en las mujeres que en los hombres porque las mujeres desempeñan, por lo general, tareas menos interesantes o atractivas.

II) Las mejoras en los medios de transporte o los posteriores incrementos en la densidad de una población conducen a una creciente división del trabajo y a la difuminación de los tipos de organización territorial, indicar:

A) Las variables que los forman y el tipo a qué pertenecen según la relación que une entre sí a las variables dentro de cada enunciado.

RESPUESTA

I) Las variables son tres:

- Absentismo laboral, dependiente.
- Sexo (compara hombre con mujeres), independiente.
- Interés de las tareas laborales, dependiente del sexo e independientes respecto al absentismo.

Se pueden designar por Y, X y Z respectivamente.

II) Las variables son cuatro:

- Mejora en los medios de transporte, X.
- Incremento en la densidad de población, Z.
- Creciente división del trabajo, Y.
- Difuminación de los tipos de organización territorial.

Las dos primeras están unidas por la conjunción o que indica disyunción y con otras dos por el término conducen que señala influencia o determinación de las dos últimas variables por las dos primeras alternativamente. Por tanto,

las dos primeras son, según el enunciado, variables independientes alternativamente, respecto a las otras dos variables dependientes a la vez, como lo indica su unión por la conjunción Y.

B) Indicar los tipos de la relación que une a las variables independientes y dependientes entre sí.

RESPUESTA

I) Según el enunciado, existe entre el sexo, el interés de las tareas y el absentismo, una relación irreversible o asimétrica: X influye en Z y Z en Y y no al revés.

Estocástica, porque no parece válido afirmar que el sexo determina completamente el interés y éste el absentismo, sino sólo en un cierto grado o con una cierta probabilidad.

De coexistencia, porque al tratarse de una relación estocástica o estadística no se da una secuencia clara entre dichas variables, salvo en el caso del sexo.

Contingente porque no parece válido afirmar que X determina Z y Z, Y, independientemente de cualquier otro factor; de hecho es de suponer la existencia de otras variables que determinen también el interés laboral y el absentismo.

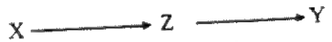
II) Asimétrica, estocástica, coexistencia y contingente por similares razones que en el caso anterior.

C) Representar mediante un modelo gráfico la relación entre dichas variables.

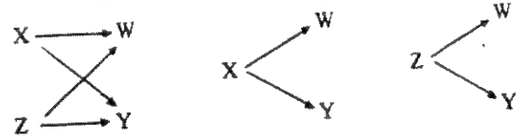
RESPUESTA

I) Según el enunciado, el absentismo laboral viene determinado por el interés de las tareas laborales, pero éste a su vez depende del sexo.

Luego el modelo que expresa gráficamente estas relaciones será el siguiente.



II) Según el enunciado, la influencia de las variables X y Z sobre Y y W es alternativa, por lo que no excluye que pueda darse a la vez. En consecuencia, puede representarse dicha influencia conjunta o separadamente:



24

Dadas las siguientes variables: edad, sexo, grado de instrucción, peso, estado civil, rango en agresividad, perímetro torácico; rasgo caracterial preferido (amabilidad, simpatía, etc.), nivel de ingresos, cociente de inteligencia, rango de prestigio, religión, grado de generosidad, número de hijos, filiación política, frecuencia de v

sitas a amigos, rango en la aptitud musical; clase social, índice cefálico; tipos de desviación social (alcoholismo, prostitución, etc.), animales domésticos; número de cabezas de ganado que se poseen; clases sociales, puntuación obtenida en una escala tipo Thurstone; temperatura; tipo de las revistas que se compran; su número; escalas de prestigio de profesiones; extensión de fincas cultivadas; deportes practicados; grados de afición al cine.

Se pide clasificarlas según el carácter de las escalas de valores a que normalmente den lugar en nominales, ordinales, de intervalo y de razón, e indicar si esta clasificación es absoluta o no.

RESPUESTA

Nominales	*Ordinales*	*De intervalo*	*De razón*
Sexo.	Grado de instrucción.	Coeficiente de inteligencia.	Edad.
Estado civil.	Rango en agresividad.	Indice cefálico.	Peso.
Rasgo caracterial preferido.	Rango de prestigio.	Puntuación obtenida en una escala, integrismo-progresismo tipo Thurstone.	Perímetro torácico.
Religión.	Grado de generosidad.	Temperatura.	Ingresos percibidos.
Tipo de desviación social.	Clases sociales.		Número de hijos.
Animales domésticos.	Grado de afición al cine.		Número de cabezas de ganado que se poseen.
Tipo de revistas compradas.	Nivel de instrucción.		Número de revistas compradas.
Deportes practicados.			

No es absoluta, porque, por ejemplo, las variables nominales se las puede dar un rango convencional: así, al estado civil: 1, soltero; 2, casado; 3, viudo, y 4, divorciado.

Igualmente las escalas de intervalo y de razón pueden constituir escalas ordinales, cuando, como por ejemplo, la edad no se usa como expresión de número de años de una persona, sino de su madurez humana o el desarrollo de su personalidad. También depende esto de las categorías que se distingan, pues si, por ejemplo, en el peso se distinguen las tres categorías siguientes: pesados, de peso medio y ligeros, entonces se trata más bien de una escala ordinal que de razón.

El cociente de inteligencia y la puntuación en la escala se suele estimar que, por la forma de su construcción tienen distancia, aunque no exista una unidad de medida en ellas.

25

Dadas las siguientes variables tomadas del estudio de A. de Miguel en Tres estudios para un sistema de indicadores sociales, Foessa, 1967, se pide distinguir en ellas, según su mayor o menor generalidad o concreción, las variables principales o generales, las variables intermedias y los indicadores según la variable general a que se refieran.

Tasa de fecundidad general; población; mortalidad; estratificación social; nivel económico; conciencia de clase; desarrollo industrial; grado de identificación de cla-

se; consumo de energía por habitante; participación política; sentimiento de futilidad de la tarea; voto en el último referéndum; vida política y asociativa; alienación; marginación y pobreza; elección del cónyuge; duración media del noviazgo; nivel de nutrición; alimentación; familia; nivel calórico por habitante; educación; urbanismo y vivienda; trabajo; edad media de terminación de los estudios; las condiciones de trabajo; proporción de viviendas compartidas por más de una familia; tiempo empleado en trasladarse al lugar de trabajo; vivienda; niveles cualitativos del sistema educativo; estratificación social.

RESPUESTA

Variables generales	*Variables intermedias*	*Indicadores*
Población	Mortalidad	Tasa de fecundidad general
Estratificación social	Conciencia de clase	Grado de identificación de clase
Nivel económico	Desarrollo industrial	Consumo de energía eléctrica por habitante
Vida política y asociativa	Participación política	Voto en el último referendum
Marginación y pobreza	Alienación	Sentimiento de futilidad de la tarea
Familia	Elección de cónyuge	Duración media del noviazgo
Alimentación	Nivel de nutrición	Nivel calórico por habitante
Educación	Niveles cualitativos del sistema educativo	Edad media de terminación de los estudios
Urbanismo y vivienda	Vivienda	Proporción de viviendas compartidas por más de una familia
Trabajo	Condiciones de trabajo	Tiempo empleado en trasladarse al lugar de trabajo.

26

Supuesto que se quiere estudiar la integración social de una comunidad, localidad de 2.000 habitantes, exponer razonadamente las dimensiones que se podrían considerar en la variable integración social, teniendo en cuenta, por ejemplo, que los aspectos fundamentales de la vida social son los individuos, los valores y las normas sociales, y que la integración implica la unión o cohesión entre los diversos elementos de un todo.

RESPUESTA

Con base en las ideas acabadas de iniciar se pueden distinguir las siguientes dimensiones de la variable integración social: a) integración cultural entre las diversas normas y valores del grupo; b) integración entre los miembros de la sociedad; c) integración de las personas respecto a las normas y valores sociales.

A su vez las relaciones personales en la sociedad tienen lugar mediante el cambio de signos y funciones. De aquí que se pueda distinguir asimismo una integración comunicativa, referida y con base en el intercambio de signos y significados, y una integración funcional, derivada de la educación de los distintos cometidos sociales.

En síntesis, las dimensiones de la integración son cinco: cultural, social, normativa, comunicativa y funcional. (Basado en W.S. Landecker: *Les types d'intégration et leur mesure*, en Le Vocabulaire des sciences sociales. Concepts et indices. París, 1957. pág. 37 y siguentes.)

27

Señalar en las siguientes variables algunas de sus categorías o valores posibles, su definición operatoria, el procedimiento para clasificar las unidades de observación en las categorías, así como los tipos a que pertenecen según su naturaleza, amplitud, nivel de abstracción y escala.

A) Edad.

RESPUESTA

Definición operativa: (hay que tener en cuenta que ésta puede ser distinta, según las circunstancias de la investigación y el grado de precisión deseado). Una de ellas podría ser:

Años ya cumplidos que se tienen en el momento de realizar la encuesta.

Categorías o valores: 18, 20, 25 años etc.

Procedimiento: Pregunta en un cuestionario, ¿Cuántos años tiene Vd ya cumplidos?

Tipos a que pertenece: por su naturaleza, es cuantitativa; por su amplitud, individual de base: por su nivel de abstracción, indicador; y por la escala que forman sus valores, de razón o proporción.

B) Nivel de educación.

RESPUESTA

Definición operativa. Grado que posee en el momento de realizar la encuesta dentro del sistema educativo nacional.

Categorías: E.G.B. incompleta. E.G.B. completa; BUP: Licenciado; etc.

Procedimiento. Pregunta en un cuestionario.

Tipos. Por su naturaleza, cualitativa. Por su amplitud, individual de base. Por su nivel de abstracción, indicador. Por su escala, ordinal.

C) Consumo de acero anual per capita.

RESPUESTA

Definición operativa. Razón o cociente entre el número anual de Kg. de acero consumidos en España y la población de cada año.

Categorías. 20 Kg, 5 Kg, 100, etc.

Procedimiento. Obtención mediante las fuentes estadísticas del consumo anual de acero en España y de la población anual.

Tipos. Por su naturaleza, cuantitativa; por su amplitud, colectiva global; por su nivel de abstracción, indicador; por su escala, de razón.

D) Religión profesada.

RESPUESTA

Definición operativa. Iglesia o confesión a la que se declara pertener en el momento de la encuesta.

Categorías. Católico. Protestantes. Ninguna, etc.

Procedimiento. Pregunta en cuestionario.

Tipos. Por su naturaleza, cualitativa. Por su amplitud, individual de base. Por su abstracción, indicador. Por su escala, nominal.

E) Conservadurismo, en el aspecto económico.

RESPUESTA

Definición operativa. Grado de conservadurismo de los encuestados, establecido mediante escala de actitudes.

Categorías. 15, 25, 40, 7, 2, etc.

Procedimiento. Aplicación de la escala en cuestión, p. ej. la R-C de Nettler tipo Likert. Tipos. Por su naturaleza, cualitativas-cuantitativas, en cuanto si bien las medidas se expresan numericamente, éstos números no representan en realidad distancias iguales sino a lo sumo un orden. Por su amplitud, individual de personalidad. Por su abstracción, la variable en sí es general, midiéndose por la serie de items o indicadores que forman la escala utilizada. Por su escala, es ordinal, por la razón indicada.

28

Dado el tema a investigar a que se refiere el ejercicio 6, es decir, el estudio de la incidencia del desarrollo económico social a partir de la posguerra en las regiones industrializadas del norte, nordeste y levante, por una parte, y el resto de España, por otra, se pide señalar las dimensiones o variables intermedias del desarrollo económico y social y señalar algunos de los indicadores.

RESPUESTA

En este caso, son múltiples y precisas, por lo general, tanto las variables intermedias que expresas aspectos de la variable general, como las variables empíricas o indicadores que miden o pueden expresar su extensión. A este respecto se pueden consultar la obra citada: "Tres estudios para un sistema de indicadores", Foessa, 1967. Aquí se señalan algunas de las variables intermedias, seguidas de varios indicadores significativos.

a) Urbanización. Indicadores: 1. Porcentaje de la población que vive en ciudades. 2. Porcentaje de la población dedicada a la agricultura.

b) Industrialización. Indicadores: 1. Porcentajes de distribución de la población laboral en las distintas categorías de industrias. Consumo de acero "per cápita".

Consumo de cemento "per cápita". Consumo de gasolina "per cápita". Consumo de energía por habitante. Tasa de productos manufacturados en las exportaciones. Número de camiones matriculados por 10.000 habitantes.

c) Nivel educativo. Indicadores: Tasa de analfabetismo. Número medio de años de escolaridad. Porcentaje de población enrolada en la enseñanza media y superior.

d) Nivel de consumo: Número de teléfonos por 1.000 habitantes. Número de automóviles matriculados por 1.000 habitantes. Porcentaje de hogares con frigorífico. Porcentaje de hogares con televisión. Porcentaje dedicado a la alimentación en los gastos totales

La adopción de estos indicadores se hallaría condicionado a la existencia de datos estadísticos sobre los mismos con referencia a las regiones estudiadas separadamente.

29

Supuesto que se ha proyectado realizar una investigación empírica para estudiar el grado de reacción social a la desviación, se pide señalar: a) si dicho grado es en este estudio variable dependiente o independiente; b) señalar las posibles variables dependientes o independientes, según el caso, que correspondan a la variable en cuestión; c) formular con dichas variables la hipótesis de trabajo; d) indicar, en fin, qué variables intermedias se podrían utilizar para el estudio proyectado.

RESPUESTA

a) El grado de reacción social a la desviación es el objeto de la investigación, lo que se pretende explicar en ella, el resultado buscado; luego se trata de una variable dependiente.

b) En la reacción social a la desviación se pueden distinguir tres tipos de sujetos: los desviados, los agentes del orden y la sociedad de que se trate. Por tanto, se pueden señalar como variables independientes o explicativas: 1) la categoría social de los desviados; 2) la posición social respecto a ellos de los agentes del orden; 3) la resonancia social de los hechos delictivos.

c) Dadas las indicadas variables, se podría formular la siguiente hipótesis: El grado de reacción social a la desviación está en relación con la categoría social de los desviados; su posición social respecto a los agentes del orden y la resonancia social de los hechos delictivos.

Dado que la influencia social de los desviados y su posición en relación a los agentes del orden puede ser mayor o menor, dos variables intermedias podrían ser: Influencia de los desviantes respecto a las fuerzas del orden y distancia social entre ellos (la escasa distancia social tiende a convencionalizar la desviación).

En cuanto a la resonancia social de la desviación, cuanto mayor sea ésta, tanto mayor será la presión social para intervenir. Por ello se puede señalar como una tercera variable. La visibilidad social de la desviación. Por otra parte, la resonancia social está también sin duda en relación con el número y calidad de las víctimas. Otra variable, pues, sería la victimología percibida de la desviación. En fin, parece que debe influir en la reacción social en cuestión una quinta variable, la racionalidad o irracionalidad percibida de la desviación. Por ejemplo, actos debidos a extrema necesidad, debilidad mental o, por el contrario, gratuitos, sin motivo alguno. (Basado

en David Downes y Pau Rock. Social reaction to deviance an its effects on crimen and criminal careers. *The British Journal of Sociology*, 22, 4, 1971).

29 B

Dada la variable participación asociativa no política, se pide determinar las variables intermedias y los indicadores de éstas, y construir para cada variable intermedia índices asignando a sus indicadores el valor relativo que se estime apropiado.

RESPUESTA

Como dimensiones de la participación asociativa no política se pueden señalar las siguientes:

a) Participación en asociaciones voluntarias; b) participación en asociaciones religiosas; c) participación comunitaria; d) interacción amistosa; d) interacción con vecinos.

Determinación de indicadores y construcción de índices. Para evitar repeticiones se señalan los indicadores atribuyendo, a título de ejemplo, un valor arbitrario a cada uno dentro de cada variable intermedia.

En la primera variable, por cada asociación voluntaria a que una persona pertenezca se le podría asignar un punto; por la asistencia por lo menos a la mitad de las reuniones, dos por asociación, y por desempeñar cargos públicos o formar parte de la directiva, tres.

En la segunda variable se procedería de manera similar, asignando a la pertenencia a una iglesia dos puntos; a la asistencia frecuente a los servicios religiosos, otros dos; a la pertenencia a asociaciones religiosas, un punto; a la asistencia a más de la mitad de las reuniones, dos, y a la ocupación de puestos directivos en ella, tres.

Respecto a la tercera variable, participación comunitaria, se podrían dar dos puntos por la participación frecuente en actos comunitarios, y cinco por participar en obras y proyectos de servicio o asistencia pública.

En cuanto a la interacción amistosa, se podrían asignar los siguientes puntos a cada indicador: Tenencia de amigos, un punto; frecuencia mínima mensual de contactos con ellos, otro punto, y dos si fuera superior.

En cuanto a los vecinos, se podría conceder un punto por tratar de tú a tres, y dos, a más de tres; y por la frecuencia semanal de hablar con ellos, un punto, y diaria, dos.

(Basado en M.E. Olsen. Social participation and voting tournot. American Sociological Review.)

30

Supuesto que se ha proyectado realizar una encuesta a familias pobres de Madrid para conocer sus condiciones objetivas de vida y sus opiniones expectativas y actitudes subjetivas en relación a su situación de pobreza, se pide determinar las variables

respecto a las cuales recoger información, con distinción de las fundamentales, las principales en relación al objeto de la investigación y otras especiales que se estime puedan ser de interés

RESPUESTA

Variables fundamentales (background).

Sexo, edad, estado civil, estudios realizados, lugar de nacimiento, su número de habitantes, profesión o especialidad poseída, años de vida en Madrid.

Variables principales:

Ocupación o trabajo principal actual.

Número de veces que ha estado parado.

Ingresos obtenidos por el trabajo principal.

Ingresos totales familiares.

Seguros que les amparan.

Ayudas económicas recibidas y pedidas.

Juego de lotería y similares.

Medios de distracción.

Empleo del ocio.

Pertenencia a asociaciones.

Condiciones de la vivienda.

Opiniones sobre el matrimonio, familia e hijos.

Opiniones sobre necesidades propias y de los pobres.

Aspiraciones familiares.

Aspiraciones profesionales.

Aspiraciones en relación a los hijos.

Expectativas respecto a la ancianidad.

Opiniones sobre su situación de pobreza.

Opiniones sobre la riqueza y pobreza en el país.

Opiniones sobre el socorro de los pobres.

Conformidad con la pobreza.

Percepción del derecho de ayuda por parte de los ricos.

Variables especiales:

Desgracias graves habidas en la familia.

Familiares próximos internados en hospital, asilo, penal, etc.

Posición económica de los padres.

Posición económica cinco años antes.

Clase social subjetiva.

(Basado en el cuestionario que figura en la obra de Demetrio Casado. Introducción a la sociología de la pobreza. Madrid, Euramérica, 1971. pág. 267.)

7. EL DISEÑO DE LA INVESTIGACION SOCIAL

NOCION

Analizadas en los capítulos anteriores la elección y determinación del problema de investigación, la formulación de hipótesis y los elementos que comprenden: unidades de observación y variables, es ahora el momento de abordar el estudio del diseño.

En el diseño se trata de planificar el contraste de las hipótesis con la realidad, lo que constituye la esencia misma de la investigación. Esta significación del diseño en la investigación exige que sea objeto de un estudio relativamente detenido.

En un sentido estricto, el significado del diseño de la investigación no se debe confundir con el de proyecto, aunque a veces estos términos se emplean indistintamente. El proyecto es un término más amplio que engloba al diseño; se puede afirmar que consiste propiamente en la especificación organizativa, temporal y económica de los distintos elementos, fases y operaciones del proceso de investigación con relación a un caso concreto de investigación.

En cambio, el diseño, de acuerdo con la acepción lingüística común del término de traza, bosquejo, esbozo de algo, se puede definir específicamente como la concepción de la forma de realizar la prueba que supone toda investigación científica y social, tanto en el aspecto de la disposición y enlace de los elementos que intervienen en ella como en el del plan a seguir en la obtención y tratamiento de los datos necesarios para verificarla.

IMPORTANCIA Y REQUISITOS DEL DISEÑO

El diseño hace referencia, por una parte, a la esencia de la investigación social, la prueba y la aplicación del método científico y, por otra,

a la esencia del trabajo científico, la validez. De aquí su importancia básica, reconocida ya en campos tales como la Arquitectura y en la Ingeniería, donde constituye, sin duda, una disciplina fundamental. La construcción de edificios y obras y la fabricación de todo nuevo producto debe responder a un diseño inteligente que combine adecuadamente y de la manera más funcional posible sus distintos elementos; planifique el desarrollo de su realización y prevea los problemas de validez y seguridad. Es claro que lo mismo se puede afirmar en el caso de la investigación científica, en general, sea o no aplicada.

La consecuencia de esta importancia básica del diseño, es la necesidad de que se le conceda la atención que merece en el estudio y formación de los científicos e investigadores. Como escribe Spector (1982, 10) "el individuo que quiere llegar a ser un investigador consumado, debe convertirse en un experto en diseño", e igualmente, continúa, quien "estudie los resultados de las investigaciones de otros, debería conocer los principios del diseño para ser capaz de emitir juicios razonables fundamentados sobre dichas investigaciones."

Entre los requisitos o condiciones de los que dependen el dominio de la técnica del diseño, se pueden destacar las siguientes:

- Un conocimiento profundo de la investigación, del método científico y de los factores que pueden afectar a su validez.
- Realismo para saberse adaptar, de modo creativo y original, a las cambiantes circunstancias de cada caso, en cuanto que, de acuerdo con Suchman (1967, 307), las decisiones del diseño deben "representar un compromiso de actuación entre las demandas rigurosas del método científico y las condiciones reales de los recursos disponibles y las exigencias de la situación investigada";
- Imaginación para intuir todos los factores que pueden afectar a la validez de la investigación de que se trata y para saber encontrar soluciones eficaces a los mismos; y
- Flexibilidad, o capacidad de adaptación del diseño inicial a las nuevas exigencias que el desarrollo de la investigación ponga de manifiesto.

El fin del diseño, como el de la investigación en general, es lograr la máxima validez posible, es decir, la correspondencia más ajustada de los resultados del estudio con la realidad. De ahí que la exposición del contenido y de los tipos de diseño, que aun falta por realizar, reclame el análisis previo de la idea de validez y de los conceptos que implica.

LA VALIDEZ Y SUS ASPECTOS

Validez, calidad de válido; válido, lo que tiene valor, y valor, se derivan etimológicamente del verbo latino *valere*, que significa estar sano, ser fuerte, poder, prevalecer. El valor y la validez se refieren, pues, a la cualidad de algo de ser sano, fuerte, potente, y, por extensión, a la bondad, la belleza y, en general, a toda cualidad positiva.

Parece obvio que en cuanto el conocimiento científico pretende conocer la realidad lo más exactamente posible, su valor o validez, su carácter positivo, consistirá en su capacidad para mostrarnos o representar esa realidad a la que se refiere.

En efecto, la investigación científica, como se ha indicado, pretende conceptualizar la realidad, es decir, obtener conocimientos o ideas que sean expresión lo más exacta posible de aquella.

Tanto por la diversa naturaleza de la realidad y de las ideas como por las limitaciones del conocimiento humano, estas ideas obtenidas en la investigación científica a lo sumo pueden llegar a ser una imagen aproximada de la realidad, pero nunca una expresión perfecta de la misma.

De ahí que se planteen en todo conocimiento humano, y de modo particular en el científico, las cuestiones siguientes:

a) Tomar las medidas necesarias y disponer las cosas de tal modo que los resultados de la investigación sean válidos y no estén viciados o adulterados por la intervención de factores extraños a la investigación, y

b) Comprobar, una vez realizada la investigación, el grado de exactitud con el que los resultados obtenidos representan a la realidad a la que se refieren.

Adoptar medidas para resolver la primera cuestión forma parte, como se deriva de lo indicado hasta aquí, del diseño, mientras que dar respuesta a la segunda exige efectuar una labor posterior al diseño y la investigación y sus operaciones y que suele recibir el nombre particular de verificación o evaluación, que será objeto de estudio en el capítulo siguiente.

Hay que aclarar que aquí se trata de esta validez, que se podría llamar genérica y no de otros tipos particulares de validez distintos, aunque relacionadas con éste, cual es el caso, por ejemplo, de la validez de un instrumento de observación o medida.

En la validez, se suelen distinguir tres aspectos o tipos: la interna, la externa y la de constructo. La primera se refiere a la concordancia, dentro de la misma investigación, de los resultados obtenidos en ella con la realidad investigada. La segunda recibe el nombre de externa porque alude a dicha concordancia pero con la realidad de otras poblaciones o fenómenos no investigados, distintos pero similares. La tercera atañe a la representatividad tanto de los indicadores como de la muestra.

El diseño debe tener en cuenta, en cuanto a la validez interna, la posible actuación en el fenómeno investigado de variables extrañas o ajenas a las que constituyen el objeto de la investigación, y, respecto de la validez externa, también de factores que afecten, en este caso, a la representatividad de los resultados de la investigación y, en consecuencia, a la posibilidad de su generalización. La existencia conocida o probable de todos estos factores hace precisa la adopción de medidas para su neutralización o control.

CONCEPTOS CONEXOS CON EL DE VALIDEZ

De acuerdo con lo acabado de exponer, la validez implica los conceptos conexos principales siguientes: las variables extrañas o externas, la representatividad, la generalización y el control.

Variables externas.— Dentro de las variables que caracterizan o intervienen en el fenómeno investigado se pueden distinguir dos grandes grupos. El primero estaría formado por las variables, dependientes e independientes, objeto de observación y estudio en la investigación de que se trate y que, por tanto, se pueden llamar internas a aquélla, y el segundo por las demás variables que, aun quedando fuera del objeto de la investigación, son o pueden constituir factores susceptibles de influir en sus resultados. A estas variables aquí se les da el nombre de variables externas o extrañas, y en inglés el de *extraneous* variables.

Existe una estrecha relación entre las variables externas y el error. Dichas variables no son otra cosa sino los múltiples factores que producen los distintos tipos de errores que se pueden producir en la investigación.

Dado que estas variables constituyen factores que pueden influir en los resultados de la investigación y mezclar sus efectos con ellos, son capaces de explicar los resultados de un estudio al margen de sus hipótesis de partida. En consecuencia, como señala Paul C. Stern (1978, 63) "las *variables externas* autorizan *explicaciones alternativas* de un conjunto de observaciones: o bien las relaciones observadas son debidas a las variables de la hipótesis, o bien se derivan por lo menos en parte de

una variable externa". Es decir, que si es así, no se podrá llegar en la investigación a una conclusión única.

Una modalidad de estas variables que merece mención especial son las llamadas en inglés *confounding variables*, término que se puede traducir por *variables parásitas*. Sucede en las variables sujeto, orgánicas o básicas, que estas variables aparecen unidas a veces indisolublemente con otras variables. En consecuencia, si se investiga una de estas variables, por ejemplo, la raza, sin tenerlas en cuenta, es obvio que los efectos que aparezcan en la investigación como producidos por esta variable, pueden ser debido a las variables parásitas, tales como los niveles diferenciales de renta, nivel educativo, formación cultural, etc., que frecuentemente son peculiares de una raza o un grupo étnico o social determinados.

Representatividad y generalización.— La representación, el actuar en lugar de otra persona y en su nombre y no en el propio, es una práctica que tiene amplia vigencia en la vida jurídica, económica y política, e incluso, en la investigación científica y, especialmente, en la social.

Especialmente en la investigación social, es frecuente que la investigación se refiera a grandes poblaciones con millones de individuos, que es imposible en la práctica encuestar individualmente, por lo que se estudian en su representación muestras extraídas de ellas. Por otra parte, las variables objeto de investigación no se pueden observar directamente, por ser muy generales y abstractas, lo que obliga a observar en su representación variables más concretas que sean indicadores de dichas variables generales.

En la vida jurídica, la representación implica que en el acto realizado por representante, éste no actúa en su propio nombre, sino en el del representado y con los mismos efectos que si éste actuara personalmente. Similarmente, en el campo de la investigación científica, el trabajo con muestras e indicadores, en representación de poblaciones más amplias y de variables abstractas, lleva consigo que los resultados de la investigación se atribuyan a la población o a la variable general, lo que supone su extensión o GENERALIZACION.

Ahora bien, si se da el caso, siempre posible, de que la muestra o los indicadores sean deficientes y no representen adecuadamente a la población o variable general a las que se refieren, entonces la generalización resultará viciada por este mismo hecho, dado que dicho defecto de representatividad implica que la muestra o los indicadores representan también a poblaciones o variables ajenas a la investigación.

Control.— La palabra control se deriva etimológicamente (E.C. Boring) del inglés counter-roll, lista principal con la que toda lista subsiguiente debía ser contrastada y, si fuera preciso, corregida.

Se ha definido este término como "el intento de eliminar el efecto de variables extrínsecas de tal modo que pueda existir seguridad de que los efectos hallados son debidos únicamente a la variable independiente probada".

Esta definición pone de relieve la estrecha conexión del control con las variables externas. En efecto, todas estas variables, si actúan en la investigación, son capaces de producir efectos no buscados ni deseados que en los resultados se mezclan con los de las variables objeto de investigación y, por tanto, los alteran. De ahí que uno de los cometidos principales del diseño sea adoptar los procedimientos de control necesarios, físicos, estadísticos o de otro tipo, para mantener las variables constantes a lo largo del proceso de investigación o eliminar su influjo. En este sentido, la afirmación de Paul E. Spector (1981, 15) de que el control es la piedra angular del diseño experimental, se puede extender a todo diseño.

ASPECTOS DEL DISEÑO Y OPERACIONES QUE COMPRENDEN

Al primer aspecto del diseño corresponde expresar o traducir en términos y relaciones operativos, referidos concretamente a fenómenos inmediatamente observables, las hipótesis; asimismo, debe prever los factores que pueden afectar a su validez.

Por tanto, comprende las operaciones principales siguientes:

a) Especificar las variables objeto de la investigación y sus relaciones, y hacerlas operativas, si son demasiado generales.

b) Determinar qué otras variables no objeto de investigación pueden influir en los resultados, bien sean externas a la investigación o bien porque aparezcan confundidas con las variables investigadas.

c) Prever los procedimientos a utilizar para controlar estas variables de modo que se asegure su no influencia o, en su defecto, la necesidad de tener en cuenta, en la interpretación de los resultados, explicaciones alternativas basadas en su posible influjo.

En cuanto al segundo aspecto, le compete la previsión de los datos sobre las variables objeto de observación que van a ser necesarios y en

especial determinan dónde obtenerlos, cómo recogerlos y modo de tratarlos.

Como operaciones principales correspondientes al efecto, se pueden señalar las siguientes:

a) Delimitación, en el espacio y en el tiempo, de la población o grupo investigado y de las unidades que se deben considerar comprendidas en ella.

b) Decisión sobre si se va a observar la población investigada entera o sólo una muestra representativa de ella.

c) La elección de las técnicas de recogida de datos adecuadas.

d) La construcción, en su caso, del instrumento de observación y su prueba o pretest.

e) Imaginar las tablas a formar con los datos obtenidos y los tipos de análisis o de otro tipo a aplicar en ellas.

De las tres operaciones distinguidas en el primer aspecto, la primera, especificación y operativización de las variables y sus relaciones, ya ha sido estudiada al tratar de las variables. Por ello, únicamente cabe subrayar aquí respecto a ella que, normalmente, ocurrirá que las definiciones operativas tendrán un grado mayor o menor de invalidez: bien porque dejan fuera aspectos de la variable representada, bien porque implican factores no incluidos en ésta.

Sobre todo en este último caso, resultará que al observar o medir la variable definida operativamente se está también midiendo factores o variables extrañas y distintas a la variable general objeto de investigación.

La única forma de control para conseguir que esto no ocurra es buscar, si es posible, otras definiciones operativas que no midan variables extrañas distintas a la investigada. Si esto no es posible, hay que pensar y se deben ofrecer explicaciones alternativas que se basen en las variables extrañas implicadas por la definición operativa utilizada.

A continuación se tratan separadamente por su importancia, las otras dos operaciones.

DETERMINACION DE LAS VARIABLES EXTERNAS A LA INVESTIGACION Y DE LAS FORMAS DE SU CONTROL

Estas variables, como se ha indicado, no forman parte de las investigadas, por tanto, quedan fuera o son externas a la investigación, pero

sin embargo constituyen factores que intervienen en el fenómeno investigado e influyen o pueden influir en los resultados obtenidos adulterándolos, de ahí la necesidad de su control.

Dada su multiplicidad y diversidad potencial, es imposible intentar siquiera una enumeración exhaustiva de las mismas, así como de las formas de control aplicables en cada caso para neutralizarlas.

No obstante, a continuación se agrupan estas variables en categorías y seguidamente se realizará una exposición de las variables externas principales que se pueden considerar incluidas en cada grupo, seguida de los procedimientos de control más comunes utilizados en cada caso.

- Ambientales.
- Conexas a las variables investigadas.
- Derivadas de la actuación del investigador en la investigación.
- Derivadas de su presencia en el campo de investigación.
- Derivadas de la forma de actuación y respuesta de los sujetos investigados.
- Ligadas al tiempo, cuando el diseño exige la repetición de las observaciones.

Variables ambientales

Son las que caracterizan al ambiente o contexto dentro del que está inmerso el fenómeno investigado, que pueden ser:

- De tipo físico: iluminación, temperatura, etc.
- De tipo social.

Son tan variadas como variados son los rasgos posibles de los contextos físicos o sociales en los que tienen lugar las investigaciones sociales concretas. En los estudios de tipo experimental, su control se realiza efectuando las observaciones en un laboratorio, lo que permite mantener iguales y constantes dichas variables a lo largo del proceso o procesos de observación que se efectúen.

En el caso de investigaciones de campo, realizadas en ambientes físicos o sociales naturales, es imposible el control directo de estas variables, pero se puede intentar determinar sus efectos mediante la aplicación de alguna de las técnicas de análisis contextual existentes.

Variables conexas a las variables investigadas

Merecen especial consideración dentro de este grupo las variables básicas, también llamadas variables sujeto y orgánicas, que siempre acompañan a los individuos y traen consigo a la investigación, sin que se puedan despojar de ellas, tales como el sexo, la edad, el estado civil, la clase social, la religión, las simpatías políticas, etc. El problema que plantean estas variables es que cuando se investiga una variable sujeto, orgánica, esta variable con frecuencia está correlacionada fuertemente con otras variables orgánicas extrañas a la investigación, por lo que los efectos observados, que se creen producidos por la variable orgánica estudiada, también serán debidos en parte y podrán ser explicados por estas otras variables.

Un procedimiento de control de estas variables es mantener constantes las variables extrañas que se crean principales a efectos de la investigación, es decir, que los sujetos investigados sean sólo, por ejemplo, hombres o mujeres, con lo que resultará controlado el sexo. Esta forma de control también se puede realizar mediante la aplicación de diversos tipos de análisis, tales como el análisis multivariable de Lazarsfeld, la correlación parcial y el análisis de covarianza.

Variables derivadas de la actuación del investigador

Aquí entra la llamada ecuación personal o conjunto de rasgos de los observadores, en cuanto pueden influir selectivamente en la percepción de la realidad observada, los cuales orientan la clase de hechos observados y su conceptualización. Son factores tales como localización o punto de vista, capacidad perceptiva, rasgos sociales, tipo de mentalidad y, de modo especial, las escuelas doctrinales o teorías a las que pertenece o se adhiere el investigador y las hipótesis por las que apuesta o pretende comprobar. Estas hipótesis, admitidas firmemente de antemano, pueden inducir a aceptar fácilmente lo que parezca confirmar la hipótesis y descuidar lo que se oponga a ellas, dando lugar a lo que Stern (1979) llama distorsión del investigador.

Todos estos factores son inherentes a la persona del investigador y, por tanto, su influjo no se puede controlar mediante su eliminación. La única solución es que el investigador explicite en su informe los rasgos físicos, psíquicos, intelectuales y sociales que le caracterizan y determinan su visión y enfoque de la realidad observada, para que se puedan juzgar la orientación de la investigación y las conclusiones obtenidas.

En cuanto a la distorsión del investigador, una forma posible de control consiste en que la apreciación o evaluación de los distintos experimentos se realice por profesionales ajenos a la investigación y mejor si lo hacen de manera ciega, es decir, sin conocer de antemano las hipótesis de partida e incluso engañados en cuanto al tenor de éstas.

Variables derivadas de la presencia del investigador en el campo de la investigación

La mera presencia del investigador en el campo de observación puede constituir también una variable externa, que influya en los resultados. Este influjo puede derivarse, en primer lugar, de la modificación de la situación observada que supone la presencia del investigador en el ámbito del fenómeno observado. Esta presencia, sobre todo respecto a grupos sociales definidos, supone la introducción de un elemento extraño que puede alterar el comportamiento habitual de los miembros del grupo.

También puede dar lugar, si la presencia es prolongada, al establecimiento de relaciones sociales diferenciales de simpatía o antipatía del observador con los distintos miembros del grupo; incluso puede producirse una cierta identificación del observador con el grupo observado, que naturalmente impedirá el relativo distanciamiento emocional necesario para observar y juzgar objetivamente su comportamiento.

Un procedimiento de control de estos efectos puede ser utilizar, cuando sea posible, observadores ocultos, no vistos ni conocidos por los investigados, o, cuando esto no sea posible, sustituir al investigador por instrucciones escritas o mediante cintas grabadas o sistemas de megafonía. También se puede comparar grupos en los que los investigadores y sus relaciones con los sujetos sean diferentes.

Son conocidas y están comprobadas, respecto a este punto, las diferencias significativas en los resultados que se obtienen en las encuestas por entrevista, cuando varía la clase social, la raza, el sexo, etc. de los entrevistadores.

Variables derivadas de la forma de actuación y de respuesta de los sujetos investigados

Dentro de este grupo ocupan un primer lugar los factores que tienen su base en propósitos pretendidos por los investigados al margen y distintos de los que supone la investigación. Reciben a veces el nombre

común de escenario (Stern) porque implican en los entrevistados la intención de representar en la investigación un papel predeterminado.

Entre ellos destacan los siguientes:

a) *El respeto social.* El respeto social, *social desirability* en inglés, tiene lugar, por ejemplo, en los cuestionarios, cuando los encuestados disfrazan sus respuestas por no dar la impresión de que sus ideas y acciones chocan con lo que se lleva o es común en la sociedad en la que viven y por no causar una impresión desfavorable de sí mismos. Incluso pueden no conocer la respuesta a la pregunta y contestar sin indicar este defecto. Es posible un cierto control de esta tendencia mediante una cuidadosa redacción de las preguntas.

b) *La aprensión evaluativa.* En este caso, las personas investigadas piensan (Rosenberg) que los observadores tienen el cometido de juzgar o evaluar sus conocimientos, salud mental, ideología, etc., y, en consecuencia, actúan en la investigación de acuerdo con este convencimiento. A efectos de su control, se puede tratar de convencer a los sujetos de que no se pretende alcanzar fin alguno de este tipo mediante la investigación, o bien replicar el experimento alterando los aspectos sospechosos de producir este efecto y la reinterpretación de los resultados, por distintos investigadores.

c) *Características de la demanda.* Orne da este nombre a una serie de pistas que en la investigación pueden transmitir la hipótesis experimental a los sujetos. Da lugar a que su respuesta al estímulo aplicado no se derive sólo de dicho estímulo, sino también de la idea que se hacen de los objetivos de la investigación y de las hipótesis que creen se pretenden contrastar.

Una forma de control de la característica de la demanda es utilizar en la investigación un grupo de control en el que dichas características sea distinta, así como el cuasi-control, término con el que Orne (1969) designa diversas actuaciones con los grupos investigados u otros de la misma población orientadas a determinar si ha actuado o no la característica de la demanda en la investigación de que se trate.

d) *Las expectativas del investigador.* Este factor, relacionado con el anterior, se deriva (Rosenthal) de que el investigador puede comunicar sutilmente a los investigados sus expectativas respecto a la investigación, lo que desea y espera que tenga lugar, por ejemplo, confirmar la hipótesis, y los sujetos pueden acomodarse en sus respuestas a dichas expectativas.

Esta variable se puede controlar empleando técnicas de observación o medidas ciegas, en las que el observador no conoce la conducta que se espera del sujeto y, por tanto, es imposible que cree ninguna expectativa en él, lo que sucede, por ejemplo, cuando los que se ponen en contacto con los sujetos y recogen las observaciones son ajenos al equipo investigador y no conocen sus hipótesis.

Extensión de este último procedimiento es la *técnica del doble ciego*, en la que el observador y los sujetos son ciegos respecto al tratamiento aplicado. Sería el caso, por ejemplo, de la administración de preparaciones que tienen el mismo olor, sabor, forma, etc. unas con una sustancia y otras sin ella, sin que los observadores ni los sujetos observados sepan distinguirlas.

c) *Efecto Hawthorne.* Recibe este nombre porque fue descubierto en los experimentos realizados por Elton Mayo en la planta industrial de Hawthorne. El objetivo de la investigación era determinar los efectos sobre la producción de variaciones en las condiciones de trabajo: iluminación, temperatura, horario.

El resultado obtenido fue que la productividad aumentaba siempre, independientemente del sentido de la manipulación: mejor o peor iluminación, etc. De ahí que dicho aumento se atribuyera no al cambio en las condiciones de trabajo, sino al efecto derivado de la conciencia de los investigados de estar siendo tratados de modo especial y por la novedad del cambio y de la situación. Sobre todo es probable que se produzca este efecto, cuando los sujetos investigados están marginados y faltos de contacto social.

Una forma de control de este efecto, válida en general también en los restantes casos expuestos, es el empleo de procedimientos de observación o medida que no interfieran en los sujetos porque se les observa o mide sin su conocimiento. Es obvio que si los sujetos no tienen conciencia alguna del hecho de la investigación, ésta no puede producir efecto alguno en sus respuestas o comportamientos.

f) *Dependencia de la memoria.* En las investigaciones referidas al pasado, sobre sucesos y fenómenos retrospectivos, los recuerdos de los sujetos investigados respecto a los mismos no sólo pueden ser incompletos, sino sistemáticamente incompletos, debido a olvidos y recuerdos inexactos. Cuando se trata de investigación por encuesta, una forma de control de este factor es redactar las preguntas de manera que ayude y facilite el recuerdo, mediante su división en varias fases escalonadas temporalmente a partir de los períodos más cercanos, o bien, su referencia a períodos de tiempo específicos y no genéricos y a notas y recibos

que eviten el "telescoping" o el recuerdo de un suceso como sucedido antes o después de lo que fue realmente. Por otra parte, los datos de este tipo se pueden contrastar también utilizando fuentes de información complementarias, si existen.

Variables ligadas al tiempo, cuando el diseño exige la repetición de las observaciones

Estos efectos son típicos de las investigaciones de tipo experimental en las que existen dos medidas del grupo o grupos investigados, una antes de que la variable independiente haya podido ejercer su influjo y otra después.

Han sido estudiados de modo especial por Campbell y Stanley (1973) y se enumeran y definen, teniendo en cuenta a dichos autores, como sigue:

Historia. Acontecimientos ocurridos entre la primera y la segunda medición, además de la variable experimental. Es obvio que, cuando son conocidos por los sujetos, pueden producir cambios en ellos, que se añaden a los causados por la variable experimental.

Maduración. Variaciones internas de los sujetos experimentados ocurridos entre el pre- y el postest, tales como el hambre, la fatiga y similares. Se trata de procesos biológicos y psicológicos que producen cambios en los sujetos y pueden alterar su respuesta a la variable experimental.

Administración del test. Influjo que la aplicación de un instrumento de medida: escala, test, cuestionario, puede ejercer sobre la segunda aplicación, como, por ejemplo, un entrenamiento que de lugar a puntuaciones más elevadas. De hecho, en las pruebas de inteligencia, se ha observado una mejora de 3 a 5 puntos en el C.I. de los sujetos a quienes se había aplicado antes el mismo test.

Instrumentación. Cambios en los instrumentos de medida, o en los observadores participantes, sufridos de una a otra medición, que pueden hacer variar las diversas mediciones que se realicen.

Selección muestral. Aquí entra la desigualdad de los grupos que intervienen en el experimento, derivada de una elección sesgada de sus muestras. Este factor puede interactuar con la maduración y la misma variación experimental.

Regresión estadística. Opera en los grupos seleccionados sobre la base de sus valores extremos. En este caso, se ha observado una tendencia a la regresión a posiciones más moderadas en las sucesivas mediciones. Se considera debida a la correlación imperfecta existente entre las medidas del pretest y del postest.

Mortalidad experimental. Se refiere a la desaparición de una u otra observación por muerte, traslado, negativa... de elementos de los grupos experimentados.

Efecto reactivo o de interacción de las pruebas. Se produce cuando un pretest aumenta o disminuye la sensibilidad de los sujetos del experimento al influjo de la variable experimental. Un pretest para medir la actitud antirracista, puede ocasionar que la reacción de los sujetos de un grupo al estímulo de una película antirracista sea más o menos fuerte que lo normal en su población.

Los efectos reactivos de la situación experimental en los sujetos al experimento. La artificialidad de la situación experimental puede impedir generalizar el efecto de la variable experimental observado en un grupo determinado a las personas expuestas a ella en una situación no experimental.

Procedimientos comunes para controlar estas variables externas son los siguientes:

Uso de un grupo de comparación o de control. Este grupo no se somete al estímulo de la variable independiente, sino a un placebo o tratamiento neutral.

Diseño ABA. En él se toma primero una "baseline", es decir se observa a los sujetos antes del tratamiento, situación A. Después se aplica el tratamiento, situación B. Finalmente se observa nuevamente a los sujetos sin someterlos a tratamiento, situación A. Este diseño controla las variables ligadas al tiempo, porque éstas se manifestarán haya o no tratamiento.

Aleatorización. Consiste en la elección al azar de los sujetos asignados a cada grupo, con lo que se evitará probablemente que los sujetos difieran sistemáticamente en términos de las variables básicas.

Contrabalanceo. En esta forma de control de los efectos de órdenes distintos en la aplicación de tratamientos diferentes, se aplica a cada grupo experimentado todos los tratamientos pero en orden diferente.

ANALISIS DE LAS OPERACIONES DEL SEGUNDO ASPECTO DEL DISEÑO: DESARROLLOS DE LA PRUEBA

Delimitación en el espacio y en el tiempo del fenómeno investigado y definición de la población o conjunto de clases de unidades de observación que abarca. La finalidad de esta operación es concretar el sector de la realidad que se va a investigar y del cual se van a obtener los datos necesarios para la prueba que supone la investigación.

Un sector de la realidad, un fenómeno científico social a estudiar, resulta especificado por tres coordenadas: el espacio en que se producen, el tiempo en que tienen lugar y el conjunto de unidades de observación o población que comprenden. Por ello, es tarea del diseño de la investigación fijar los límites espaciales o geográficos del fenómeno a estudiar y definir la población que abarca.

Aunque esta operación no siempre planteará problemas graves, es imprescindible realizarla con toda seriedad, delimitando con precisión, incluso gráficamente, con el mapa o plano oportuno, los límites geográficos del estudio, especificando igualmente, con exactitud, sus límites temporales y concretando los tipos de unidades que se consideran incluidos en la población de la cual se van a obtener los datos. Al mismo tiempo se dará una solución a los casos raros y especiales que frecuentemente se plantean, sobre todo cuando se conocen bien el campo y la población investigada.

Es imprescindible realizar esta operación, porque es la condición para que el equipo de investigación pueda ajustarse estrictamente a lo largo de todo el proceso de investigación a los límites fijados: espaciales, temporales y poblacionales y a fin de que los datos obtenidos sean homogéneos y se refieran al fenómeno estudiado definido.

Diseño de la muestra. Cuando en una investigación resulta prácticamente difícil la observación de todas las unidades que integran la población investigada, hay que limitar dicha observación a una parte o muestra del universo en cuestión. Entonces, la definición de la población plantea los problemas adicionales del cálculo del tamaño de la muestra y de su error y de la determinación de un procedimiento de elección viable, que ofrezca garantías suficientes de que la muestra es representativa de la población y es válida para contrastar las hipótesis de partida.

Dados los diversos tipos de investigaciones, la multiplicidad de tipos de poblaciones y los distintos procedimientos de elección de la muestra existentes, el diseño de la muestra requiere un estudio especial de los mismos con relación al caso concreto investigado, con el fin de establecer las características de la muestra más adecuadas a los objetivos pre-

tendidos en la investigación de que se trate, y, de modo especial, para obviar los sesgos que la utilización de una muestra no representativa puede suponer.

Se considera que una muestra no es representativa, cuando está afectada por un error sistemático que hace que determinados sectores estén super o sobrepresentados en la misma. Por ello, si sus respuestas difieren de las propias de los restantes sectores dará lugar a que los resultados obtenidos serán en parte inválidos y no generalizables por completo a la población investigada.

El procedimiento más seguro para controlar este sesgo es utilizar una muestra verdaderamente representativa. Si ello no es posible, en la interpretación de los resultados se deberá tener en cuenta el sesgo muestral y ofrecer, basados en él, explicaciones alternativas.

Elección de las técnicas de observación, construcción del instrumento de observación, su prueba y previsión de las tablas de datos precisas y de los tipos de análisis adecuados. El diseño no sólo ha de especificar la población fuente de los datos necesarios en la investigación, sino que ha de prever y planificar también cómo recoger los datos y el modo de tratarlos, una vez recogidos. A estos fines tienden las operaciones enumeradas en el epígrafe, reunidas en él a efectos de su exposición conjunta.

La elección de las técnicas de observación es una operación que se debe realizar, fundamentalmente, teniendo en cuenta las características de las unidades de observación y, sobre todo, la naturaleza de las variables empíricas, respecto a las que es necesario recoger información, así como los factores de coste y tiempo.

Es recomendable el empleo, a la vez, y si es posible, de técnicas distintas o, en su defecto, se debe idear alguna forma de contraste de los datos obtenidos. Este empleo recibe el nombre específico de triangulación. Esta se funda en que, como señalan Ch. A. Lininger y D.P. Warwick (1978, 21) "todo método de recopilación de datos, incluyendo la encuesta, constituye una aproximación al conocimiento. Cada uno proporciona una visión distinta de la realidad y todos tienen limitaciones si se utilizan en forma aislada".

La elección de las técnicas de observación reclama que se diseñe el instrumento que se empleará para la recogida de datos, determinando las características básicas que debe reunir y los elementos básicos que deben integrar su contenido, así como los procedimientos que se utilizarán para contrastar su validez y seguridad, incluida su prueba previa.

Este trabajo sólo se puede referir muy sucintamente a esta operación, como a las demás. No obstante, cabe señalar que se deberán tener en cuenta en ella, las múltiples variables externas a la investigación expuestas antes, y las ligadas al propio instrumento de observación, como, por ejemplo, en el cuestionario, las derivadas de la forma de preguntas utilizadas, de su redacción y del modo de su categorización, y adoptar los remedios que sean factibles para corregir sus efectos o neutralizarlos.

El diseño deberá también prevenir las contingencias que se pueden plantear en la aplicación del instrumento de observación elegido, que Schuman (1967) llama diseño administrativo, es decir, en la recogida de datos o trabajo de campo, y dar pautas que permitan acomodar las líneas del diseño a las circunstancias imprevistas y cambiantes de la realidad.

Por último, el diseño comprende también, en cuanto al modo de tratar los datos recogidos, la especificación de la forma de su procesamiento y de las variables que se deben cruzar o relacionar con otras en las tablas correspondientes, de acuerdo con los objetivos e hipótesis formuladas, así como las formas de análisis adecuadas a los datos recogidos y a las tablas formadas.

TIPOS DE DISEÑO Y SU REPRESENTACION

En el diseño de una investigación se pueden distinguir los siguientes elementos:

- Los sujetos o grupos y la forma de su elección.
- La variable o variables independientes o factores, cuya influencia sobre las variables dependientes observadas se estudia en la investigación; y dentro de ellas, su número de niveles o categorías.
- Las observaciones a efectuar, su carácter y el orden de realización de las mismas.
- La forma de elección de los individuos que componen los grupos investigados.

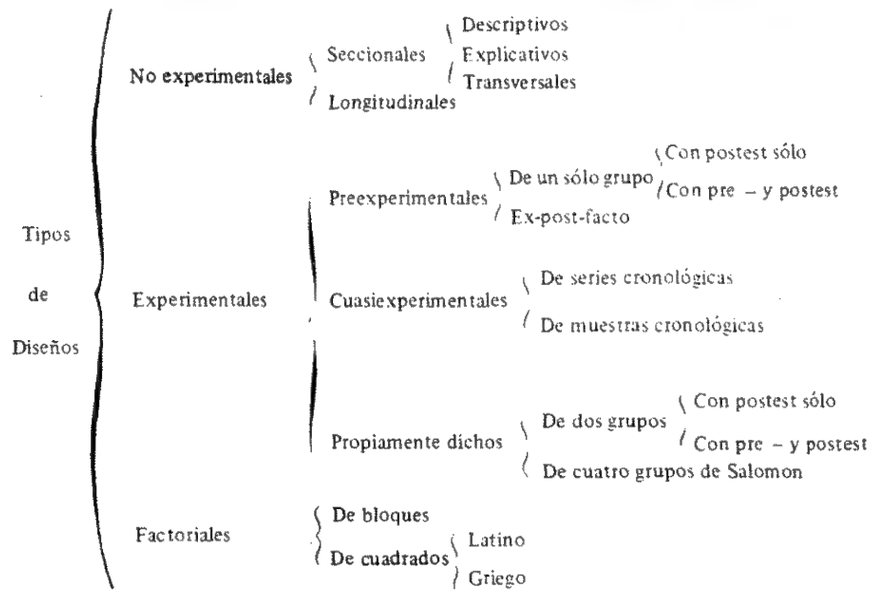
Teniendo en cuenta que el número y modalidades de estos elementos pueden ser uno o varios y dado que cada combinación distinta de elementos, su número y modalidades da lugar a un diseño distinto, hay que concluir que los tipos posibles de diseños son prácticamente innumerables, y que se trata de una cuestión no cerrada sino siempre abierta a la imaginación de los investigadores.

Por ello, aquí me limitaré a considerar algunos de los diseños más conocidos y utilizados, de acuerdo con la siguiente clasificación.

Según se refieran a estudios no experimentales o experimentales, los diseños se pueden dividir, en primer lugar, en diseños no experimentales y experimentales.

Los estudios no experimentales pueden consistir en una sola observación, o en varias, unas después de otras, dando lugar respectivamente a los diseños seccionales y longitudinales, respectivamente. A su vez, los diseños seccionales pueden ser descriptivos, explicativos y transversales.

En cuanto a los diseños experimentales, Campbel y Stanley (1978) distinguen, teniendo en cuenta el grado de su perfección experimental, los diseños preexperimentales, cuasiexperimentales y experimentales propiamente dichos, y, dentro de ellos, diversas clases de diseños entre los que cabe destacar los que se consignan en el esquema siguiente.



Además existen los diseños factoriales, con dos o más variables independientes, en los que se pueden separar las modalidades de individuales, en bloques o grupos y de los cuadrados latino y griego.

La representación de los diseños se basa en la notación mediante letras y números de sus elementos básicos. En el sistema seguido por

Campbell y Stanley, los grupos, si son varios, se indican asignando una línea distinta a cada grupo; la variable experimental, con una X y subíndices, si son varias; las observaciones o mediciones, con una O y subíndices numéricos correlativos, si son varias; finalmente, la elección aleatorizada de los grupos se indica mediante una R, pero, como dicha elección también se puede efectuar mediante igualación, en este caso, se podría emplear una I.

En lugar de designar las observaciones por O y las variables experimentales por X, se representa también por A la observación, *baseline*, de un grupo sin actuación de la variable experimental y por B, con actuación de dicha variable.

DISEÑOS NO EXPERIMENTALES

Diseños seccionales

De los distintos elementos que pueden intervenir en los diseños, los seccionales no comprenden ni diversidad de observaciones, ni de grupos, ni tampoco variables experimentales; quedan limitados a una sola observación de un solo grupo en un solo momento del tiempo. Por eso son los diseños más simples posibles de tal modo que se representan mediante una O, a la que se podría añadir antes una R, en el caso que se quiera indicar que el grupo observado es una muestra elegida aleatoriamente. R O designa una única observación de un solo grupo elegido aleatoriamente.

A pesar de su simplicidad o, quizás por ella misma, esta clase de diseño es la más frecuente en las investigaciones sociales. Emplea técnicas de recogida de datos basadas en la observación directa, participante o no, en la encuesta y el análisis de documentos.

Tienen la ventaja de que se basan en la observación de los objetos de investigación tal como existen en la realidad, sin intervenir en ellos ni manipularlos. Esta ventaja implica también, por otra parte, su limitación, pues mediante estos diseños, en principio, sólo se pueden estudiar los rasgos de los fenómenos pero no sus causas y efectos. No obstante lo anterior, este diseño, sobre todo en el caso del empleo de técnicas de encuesta, permite observar a la vez muchas variables de los individuos que forman los grupos observados, incluso referentes a períodos de tiempo distintos, y, posteriormente, aplicar el análisis estadístico para hallar correlaciones entre ellas e incluso formar modelos de relaciones entre las variables, causales o de otro tipo, y contrastar después su ajuste a los datos, mediante las diversas técnicas de análisis multivariable actualmente existentes.

Estos diseños están sujetos, en principio, a todas las variables externas indicadas derivadas de la actuación del investigador y de su presencia en el campo de la investigación, así como al efecto de las variables de escenario, de las vinculadas a la memoria y de las parásitas, pero no así a las ligadas al transcurso del tiempo. Por tanto, son aplicables en cada caso, las formas de control señaladas antes al tratar de estas variables externas. También pueden estar sometidas al influjo de variables externas peculiares, derivadas, por ejemplo, en el caso de la encuesta, de la forma y redacción de las preguntas y su orden en el cuestionario.

El diseño seccional descriptivo tiene lugar cuando se estudia descriptivamente un grupo social en un momento dado. Por ejemplo, análisis del fracaso escolar en una comunidad estudiantil con distinción del nivel y tipo de estudios, del sexo de los alumnos, de las clases de centros de enseñanza, de la ocupación de los padres, de su clase social, etc.

El diseño explicativo pretende inspirarse en el modelo de los diseños experimentales. En él, la investigación se centra en la comparación de grupos que difieren en términos de variables determinadas, que se consideran como independientes, a la vez que se procura el control de otras variables mediante la preparación de grupos homogéneos: por ejemplo de igual edad, ingresos, clase social, etc. Se orienta normalmente al contraste de hipótesis determinadas, lo que exige la preparación de muestras específicas formadas por los subgrupos diversos exigidos por dicho contraste.

El diseño seccional-transversal *(cross-sectional* en inglés) se produce cuando se distinguen en la población investigada distintos grupos de edad con el fin de inferir la evolución con el paso del tiempo de la variable o variables observadas. Por ejemplo, investigar en un momento determinado el tiempo de irse a la cama de niños de diferentes edades, con el propósito de inferir cómo evoluciona con el paso del tiempo la hora de acostarse de los niños, sin necesidad de tomar un grupo de niños de igual edad y repetir a lo largo del tiempo las observaciones de su hora de acostarse.

Diseño longitudinal

Estos diseños se diferencian de los seccionales en que en lugar de constar de una sola observación, comprenden dos o más medidas realizadas al mismo grupo en tiempos diversos. Se representan por tantas O como números de medida repetidas se realicen, distinguiéndolas con un subíndice que indique su numeración correlativa: O_1 O_2 ... O_n.

Aunque no tanto como el seccional, también es frecuente en las investigaciones sociales, y, sobre todo, en los trabajos secundarios que utilizan series anuales de datos estadísticos.

Presentan también similares ventajas e inconvenientes que los seccionales, si bien la repetición de observaciones permite estudiar la evolución a lo largo del tiempo de las variables medidas, e incluso puede proporcionar una base mayor para análisis multivariables de causalidad. Este diseño está expuesto a todas las variables externas, incluso a las ligadas al paso del tiempo de una a otra observación.

DISEÑOS EXPERIMENTALES

Diseño de un grupo experimental con una sola medición

Consta de una observación de un grupo, en el que se ha hecho incidir previamente un estímulo o variable experimental. Su esquema, por tanto, es el siguiente: X O.

Participa del experimento en cuanto se da una manipulación del grupo observado. Sin embargo, como señalan Campbell y Stanley (1978, 19) "adolecen de tan absoluta falta de control que su valor científico es casi nulo", de ahí que dichos autores lo califiquen de preexperimental. Tampoco permiten la comparación, básica en la investigación experimental, salvo que se realice con otras observaciones distintas recordadas casualmente o con simples previsiones de lo que hubiera sucedido de no haber actuado el estímulo.

Diseño de un solo grupo con pretest y postest

Se diferencia del anterior, en que consta no sólo de postest, sino también de un pretest o medida del grupo antes de su tratamiento con la variable experimental.

Su esquema es: O_1 X O_2

En él, por tanto, existe una situación base A, en inglés *baseline*, o punto de comparación: la medida inicial del grupo, antes de sufrir el impacto del estímulo aplicado o situación B. De ahí que se le represente y reconozca también como diseño A B.

No obstante esta mejora, persiste en él, en gran parte, la falta de control de las variables externas, especialmente de las ligadas al paso del tiempo de una a otra observación, por lo que lo consideran también preexperimental.

Se basa en el supuesto de que la variación de una a otra medida se debe al influjo de la variable experimental, pero no da seguridad alguna de que dicha variación observada no sea debida, en todo o en parte, a los diversos tipos de variables externas al fenómeno. Por ello sólo permite conclusiones tentativas.

Tampoco informa de lo que sucede cuando deja de actuar la variable experimental. Para conocer este extremo, es preciso ampliar este diseño con una nueva medida del grupo sin influjo del estímulo, tipo A B A. En este caso, sin que se llegue al control de las variables externas ligadas al tiempo, cuando se puede suponer que el influjo de éstas es reducido o no existe, se puede concluir con un elevado grado de seguridad que el estímulo es responsable de los cambios observados en B.

Este diseño se mejora, aún más, si se añade otra situación B, dando lugar al diseño A B A B, el cual proporciona dos ocasiones para comprobar el efecto de la variable independiente.

Diseño ex-post-facto.— Este diseño se produce cuando falta el control de la situación inicial y el del estímulo, ya que se realiza después de haber actuado la variable experimental. En él se comparan dos grupos igualados por el investigador: uno que ha sufrido un cierto impacto y otro no. Su esquema es el siguiente:

X O I
O I

La I indica la igualación posterior de los grupos.

Su validez interna y externa es muy deficiente; sin embargo son objeto de una cierta aplicación en las investigaciones sociales, como lo demuestra la obra de Greenwood (1951).

Diseño de series cronológicas.— Este diseño es también un desarrollo del acabado de exponer, como éste era del que le precede. En lugar de una sola observación A, sin estímulo, antes y después de la situación B, o de la actuación de la variable experimental, comprende dos o más medidas repetidas de dicho tipo A A A B A A A, o bien, según la notación de Campbell:

$$O_1 \; O_2 \; O_3 \; X \; O_4 \; O_5 \; O_6 \; O_7$$

La repetición de medidas antes y después del test, proporciona un mayor poder de aislamiento de los efectos de la variable experimental, y permite controlar las variables externas: maduración, administración

del test, regresión, selección muestral y mortalidad, pero no así la historia, la instrumentación y la interacción de las pruebas, por lo que su validez interna es casi aceptable, pero no así su validez externa. La no existencia en él de grupo de control hace que Campbell y Stanley lo consideran cuasiexperimental.

Diseño de muestras cronológicas

Se trata de una derivación del anterior, a la que Campbell y Stanley dan el nombre indicado. Se caracteriza por estar formado por observacines repetidas tanto tipo A, igual que el anterior, como tipo B, derivadas éstas de la introducción reiterada de la variable experimental. Se esquematiza como sigue:

$$0_1 \quad X0_2 \quad 0_3 \quad X0_4 \quad 0_5 \quad X0_6$$

Mejora la validez interna del diseño anterior, pues permite controlar la historia. Por otra parte, admite múltiples modalidades tanto por las diversas combinaciones posibles de situaciones A y B, como por la repetición o no seguida de las mismas y la variación admitida en ellas de los niveles de la variable experimental.

En general, según se deriva de todo lo expuesto hasta aquí, estos diseños referentes a un solo grupo no consisten en otra cosa sino en combinaciones diversas de situaciones A y B. Son básicos en la investigación psicológica para el estudio de individuos o casos singulares, en los que no existe posibilidad de formar grupos de control, como lo demuestra la obra de Michel Hersen y David H. Barlow (1983). La ausencia de grupo de control se compensa en ellos, con la facilidad que permiten normalmente de practicar, respecto a cada situación A y B, no una sino diversas mediciones o pruebas al sujeto investigado.

Diseño de grupo de control con pretest y postest

Este diseño comprende, además de la variable experimental, todos los elementos de la observación experimental: las medidas antes y después del test y el grupo de control al que se aplica un tratamiento cero o neutral.

Su esquema es:

$$\begin{array}{cccc} R & 0_1 & X & 0_2 \\ R & 0'_1 & & 0'_2 \end{array}$$

La fila primera corresponde al grupo experimental y la segunda al de control. El subíndice 1, indica el pretest y el 2, el postest. La R indica la igualación de los grupos por aleatorización. Según Campbell y Stanley, este diseño es el más empleado de los diseños experimentales básicos. Permite controlar las variables externas indicadas ligadas al tiempo, excepto la interacción entre la administración del test y la variable experimental, por lo que su validez interna es aceptable pero no tanto la externa.

Diseño con grupo de control y sin pretest

Es una variante del diseño anterior en la que se suprime el pretest porque, según Campbell y Stanley, se considera que no es indispensable, ya que la elección aleatoria de los miembros de los grupos experimental y de control asegura, si es correcta, la igualdad inicial de ambos grupos.

Su esquema es el siguiente:

$$\begin{array}{ccc} R & X & 0_1 \\ R & & 0'_1 \end{array}$$

Con relación al diseño precedente, presenta la ventaja, al prescindir del pretest, de que no puede estar afectado por el factor de la administración del test ni por el de la reactividad o interacción entre el pretest y la variable experimental. Por ello, Campbell y Stanley lo consideran preferible al anterior, salvo que existan dudas fundadas de la igualdad inicial de los grupos experimental y de control o sobre la posible existencia de sesgos en la elección aleatoria de los grupos.

Diseño de cuatro grupos de Solomon

Este diseño es un desarrollo del de grupo de control con pretest y postest. Consiste en añadir a este diseño, dos nuevos grupos sin pretest: uno experimental y otro de control. Por tanto, combina los dos diseños anteriores y acumula las ventajas de ambos. Además, permite no sólo el control, sino la medida de la interacción o reactividad entre la variable experimental y el pretest. Por otra parte, la repetición de las pruebas proporciona una mayor capacidad de verificación de los efectos de la variable experimental.

Su esquema es el siguiente:

R O_1 X O_2

R O'_1 O'_2

R X O_2

R O'_2

DISEÑOS FACTORIALES

Estos diseños se distinguen por los siguientes rasgos:

1. Comprenden dos o más variables independientes y no una sola;
2. Los niveles o categorías de las variables se combinan entre sí, dando lugar a diversos tratamientos.
3. Los sujetos o grupos se asignan aleatoriamente a los distintos tratamientos.

En estos diseños se forman con los sujetos investigados igual número de grupos que de tratamientos (combinaciones posibles entre los diversos niveles de las variables independientes). Después se atribuye o aplica un tratamiento a cada uno de los grupos. Sean las variables: métodos de enseñanza, con dos niveles: método 1 y método 2; y forma de aplicación del método, también con dos niveles: audiovisual (A) y tradicional (B). Entonces, los tratamientos posibles y la representación de diseño se reflejarían en las siguientes tablas:

Método de enseñanza	Forma de aplicación	
	Tradicional, A	Audiovisual, B
1	A 1	B 1
2	A 2	B 2

R A1 01
R A2 02
R A3 03
R A4 04

Diseño de bloques aleatorizados

Estos diseños se diferencian de los factoriales en que no están formados únicamente por variables independientes, sino también por una variable extraña que se quiere controlar de los sujetos investigados, para lo cual se forman bloques distintos con los sujetos similares en los niveles o categorías de la variable extraña.

Formados los bloques o grupos, los elementos de cada uno de ellos se distribuyen aleatoriamente entre los tratamientos o bien entre los diversos niveles de la variable independiente cuando ésta es sólo una. En este último caso, la variable que se utiliza para formar los grupos se puede considerar como segunda variable.

Ejemplo: Variable independiente: 4 clases de abono: A, B, C y D. Sujetos: Parcelas divididas, para controlar la variable extraña fertilidad, en dos bloques, según su alta o baja fertilidad. Asignación aleatoria de tratamientos contrabalanceada.

Bloques	*Niveles variables ind.*			
1 (+ fértil)	A	B	C	D
2 (– fértil)	B	C	D	A

Cuadrados latino y grecolatino

Estos diseños se pueden considerar como una extensión de los diseños de bloques, pues en ellos las variables extrañas que se quieren controlar son dos. En el latino la variable independiente es una y en el grecolatino son dos. En ellos debe ser igual el número de niveles o bloques de todas las variables, de aquí su nombre de cuadrados. La asignación de tratamientos se realiza de modo que cada uno aparezca una sola vez en las filas y otras en las columnas.

Ejemplo: V. Indep.: Droga, con cuatro dosis: A, B, C y D. Sujetos: conejillos de indias, con control de las variables, de 4 niveles, la edad y el peso. Asignación de tratamientos:

Niveles de peso	*Niveles de edad* 1	2	3	4
1	A	B	C	D
2	B	C	D	A
3	C	D	A	B
4	D	A	B	C

El diseño grecolatino resulta de la conjunción de dos cuadrados latinos, de los cuales, en un caso, los tratamientos se designan con las letras del alfabeto latino y, en el otro, con las del griego. Comprende, por tanto, una variable más que el latino. Ninguno de los tipos de tratamiento se repite ni en las filas ni en las columnas.

Desarrollando el ejemplo anterior, su esquema es el siguiente:

Peso	*Edad*	1	2	3	4
	1	Aα	Bβ	Cγ	Dϵ
	2	Bβ	Cγ	Dϵ	Aα
	3	Cγ	Dϵ	Aα	Bβ
	4	Dϵ	Aα	Bβ	Cγ

EJERCICIOS

31

Supuesto que se tiene que realizar un estudio de la delincuencia infantil y juvenil en España, referida a actos libres y responsables. constitutivos de delitos e infracciones comunes en las leyes penales civiles, militares y de menores, basado en fuentes estadísticas, definir el universo y delimitar el campo de investigación en sus diversos aspectos, teniendo en cuenta que sólo se dispone de estadísticas a partir de 1954 y que no existen ficheros de delitos en todas las audiencias, pero sí en la de Madrid.

RESPUESTA

En el aspecto temporal es obvio que al basarse la investigación en fuentes estadísticas y no existir éstas antes de 1954, aquélla no podrá extenderse a años anteriores, aunque normalmente se podría partir de este año, salvo que el tiempo o los medios de que se disponga no lo permitan u otra razón lo aconsejara así.

En el aspecto geográfico también existe el inconveniente de la falta de ficheros en las Audiencias, por la dificultad que supone el manejo directo de las causas y la improbabilidad de que pudieran ser facilitadas éstas. Por ello se impondría también la necesidad de restringir la investigación a una o más de las que tengan ficheros y se estimen más representativas del conjunto.

Personal. Hay que atenerse a la clasificación legal de menores y mayores a efectos penales. Dentro de estos límites se excluiría al tratarse de delitos responsables a los delincuentes: menores y jóvenes anormales, por motivos de legítima defensa o en casos de necesidad extrema, así como los que hayan actuado por violencia o miedo incoercibles.

Objetiva. El que se refiera a los delitos comunes, excluye los delitos políticos, y los especiales, como, por ejemplo, contrabando, circulación y laborales. Aquí sería necesario un detenido examen de delitos e infracciones para ver los que se incluían.

A continuación se expone el alcance de una investigación realizada sobre este tema y publicada en la "Revista de la Opinión Pública", núm. 16, 1969.

Temporal. A partir, como se ha indicado, de 1954.

Geográfica. Se limitó a la Audiencia Territorial de Madrid.

Delimitación del universo.

Personal. Niños comprendidos de seis hasta dieciseis años de edad y jóvenes de dieciséis a veintiún años de edad sujetos a la actividad policial o de los tribunales españoles dentro de los límites temporal y geográfico indicados.

Por el objeto. Se limitó la investigación a los elementos de conducta que son reprobados explícitamente por el Código Penal español, y cuya aparición hace intervenir a los órganos mantenedores del orden.

Se excluyeron los delitos políticos, los referentes a materia de contrabando; las infracciones de tráfico que no supusieran lesiones u homicidios responsabilizados, y las infracciones laborales. Por el contrario, se incluyeron los delitos contra el Código de Justicia Militar y las infracciones sometidas a los tribunales tutelares de menores.

32 (A)

Supuesto que se tiene que realizar un estudio en España sobre las actitudes y concepciones de clase de la nueva clase obrera, se pide delimitar el campo de investigación y definir el universo.

RESPUESTA

Delimitación temporal. Del enunciado se deriva que se trata de una investigación no sobre un período de tiempo que pueda ser más o menos amplio, sino sobre la realidad actual.

Delimitación geográfica. El enunciado alude a España, pero, salvo que se disponga de medios amplios, sería difícil que pueda abarcar a todo el país. Por eso habría que restringirla a uno o varios puntos industriales del país seleccionados.

Definición del universo. Las unidades de observación en este caso son muy poco precisas, sujetos que formen la nueva clase obrera. Ello hace aún más necesario definir el universo lo más concretamente.

En primer lugar, parece claro que se debe restringir a los obreros industriales, pues en los agrícolas, aun admitiendo que se dé este fenómeno en ellos de la nueva clase obrera, presentan los fenómenos laborales características propias.

Dentro de la industria, se podría considerar constituida la nueva clase obrera por los trabajadores con una cualificación e ingresos relativamente elevados. En relación a

esta noción habría que definir concretamente dentro de qué límites se considerarían ingresos elevados, así como qué se debería estimar como especialización. Según la noción, habría que excluir a los obreros no especializados o semiespecializados, pero con ingresos altos, por cualquier circunstancia especial, así como también a los especialistas, pero con ingresos, por los motivos que fueran, reducidos. Sin embargo, podría ser conveniente por razones de comparación incluir algunos de éstos en una muestra más reducida.

Se da el caso también de los obreros independientes, especialistas, con elevados ingresos a veces, pero que trabajan por su cuenta: pintores, fontaneros, electricistas, etcétera. A éstos tampoco se les debería incluir, pues su trabajo independiente hace que no sean propiamente obreros. Otro caso especial es el de las mujeres obreras. Parece que si se dan las circunstancias indicadas de especialización y elevados ingresos, deberían ser incluidas también.

32 (B)

Supuesto que se tiene que realizar en España una investigación sobre la relación entre los conflictos laborales y los acontecimientos políticos, se pide delimitar el campo de la investigación.

RESPUESTA

Delimitación temporal. La investigación no tiene por qué restringirse en este caso al momento dado, sino que más bien debería extenderse a un período de tiempo determinado, según la amplitud que se quiera dar a la investigación y los medios disponibles: tres, seis meses, un año, etc.

Delimitación geográfica. Se puede repetir lo indicado en el ejercicio anterior: que sería difícil salvo que se disponga de un equipo y grandes medios la extensión a toda España, por lo que sería conveniente restringirla a una determinada región, e incluso dentro de ésta, a una o varias ramas industriales.

Definición del universo. Aquí se tienen dos unidades de observación: conflictos laborales y acontecimientos políticos. Sobre los primeros no se incluirían los de carácter particular, aunque afecten a más de un obrero, dado el carácter social de la investigación. Dentro de los conflictos laborales colectivos habría que comprender todos los producidos, aunque se hayan terminado pacíficamente y no hayan dado lugar a manifestaciones violentas, huelgas, boicots, sabotajes, etc.

En cuanto a los acontecimientos políticos, se consideraría como tales todos los que interesan a la opinión pública y la dividan en distintos campos ideológicos, aunque sean de carácter internacional, siempre que afecten a la comunidad nacional.

33

Supuesto que un organismo de investigación sociológica tiene en proyecto la realización de las investigaciones que se indican, se pide señalar en cada caso el procedimiento de investigación adecuado.

1) Estudio de una comunidad rural, situada en una comarca muy atrasada económica y culturalmente y sin comunicaciones con el resto de la comunidad nacional.

RESPUESTA

Este estudio es claro que habría de ser realizado mediante el traslado a esta comunidad y la recogida de los datos que interesen, empleando las técnicas de observación directa e incluso las de la observación participante y la entrevista informal.

2) Un estudio sobre la difusión de una revista, en el que una consulta previa ha puesto de manifiesto que los canales de distribución son muy deficientes, por ejemplo, que en algunos kioskos no se recibe nunca, en otros unas veces sí y otras no y en otros llega atrasada.

RESPUESTA

Se trata de un caso de los llamados estudios de mercado, que se debe llevar a efecto por el procedimiento de encuesta. No obstante, hay que advertir que en el enunciado se dice que es muy deficiente la distribución de la revista en cuestión, por lo cual es inútil realizar una encuesta sobre la aceptación por el público de la revista, mientras no se solucione esta deficiencia.

3) Un estudio descriptivo de la población obrera española emigrante que trabaja en los países centroeuropeos.

RESPUESTA

Tampoco sería necesario realizar encuesta, pues se trata de un estudio en el que existen estadísticas que se publican periódicamente, de las cuales, bien directamente o por explotación de sus datos, se podrían obtener los datos buscados. Estos datos se podrían completar con los que figuren en las solicitudes y expediente de emigración, si fuere necesario. También se debía procurar ver si existen estadísticas al respecto en los países donde están ocupados.

4) Investigar los efectos beneficiosos o perniciosos del estudio del latín y griego en la segunda enseñanza.

RESPUESTA

El medio más adecuado para llevar a cabo esta investigación sería preparar un experimento con dos grupos de alumnos, uno de estudiantes y otro de no estudiantes de latín y griego, iguales en todas las demás circunstancias, excepto en el estudio o no de dichas lenguas, es decir, manteniendo controlados o igualados todos los demás factores de diversidad.

5) Estudiar las variaciones en cantidad y calidad de la información religiosa en la prensa de 1955 a la actualidad.

6) Un estudio comparativo de las ideologías subyacentes en los textos de Historia de España de primera y segunda enseñanza desde 1920 hasta la actualidad.

RESPUESTA

Se trata, como se desprende claramente del enunciado de estudios de documentos, en el primer caso, prensa, y en el segundo, manuales de Historia de España.

7) Una investigación sobre lo que la gente piensa ser y hacer en 1985.

RESPUESTA

Habría que realizar una encuesta, pero tendría más valor como muestra de las opiniones de la gente que como predicción.

34

Dadas las siguientes hipótesis, se pide indicar en cada una de ellas las dimensiones de las variables y sus indicadores, así como señalar el método de investigación a emplear.

1) Los delincuentes comunes son irreligiosos.

RESPUESTA

En esta hipótesis la variable es la irreligiosidad. En ella, según Ch. Y. Glock (*Y a-t-il un reveil religieux aux Etats Unis?*, "Vocabulaire de Sciences Sociales, París, 1965), se pueden distinguir las cuatro dimensiones siguientes: experiencial, ritualista, ideológica y consecuencial. La primera se refiere a todos los sentimientos, percepciones y sensaciones que implican una comunicación personal con Dios. Indicadores de ella serían toda clase de fenómenos místicos: contemplación, intuiciones de lo divino, inspiraciones, actos providenciales, éxtasis, raptos, apariciones. El método normal para investigarlas sería las declaraciones del interesado por cuestionario y mejor por medio de entrevista.

Los indicadores de la dimensión ritualista son la frecuencia en el cumplimiento de las prácticas religiosas, en la recepción de sacramentos, etc. El método apropiado, la encuesta también por cuestionario o entrevista, y datos estadísticos, si los hay, en el caso de comunidades.

Los indicadores de dimensión ideológica son el grado de aceptación de los dogmas religiosos. La técnica adecuada para investigarlos sería la aplicación de escalas apropiadas.

La dimensión consecuencial se refiere a la conducta moral de los individuos. Indicadores de ella serían la práctica de las obras de misericordia, la beneficencia, el espíritu de amabilidad, servicio y solidaridad con los demás. Método a emplear, las escalas, cuestionarios y entrevistas.

2) La movilidad social descendente es más frecuente en los individuos esquizofrénicos que en los normales.

RESPUESTA

Las variables en este caso son la "movilidad social descendente" y la "salud mental", con dos dimensiones, esquizofrenia y normalidad mental. La movilidad social descendente implica la reducción en el "status social", y como los componentes de

éste básicos son los ingresos, la jerarquía y el prestigio se puede considerar como dimensiones de la movilidad social descendente la pérdida de ingresos, de jerarquía o de prestitio dentro del período examinado. Indicadores serían la variación en los ingresos percibidos, en la categoría profesional ostentada y, en cuanto al prestigio, la variación en las formas de relación y participación social en dicho período. El método apropiado sería el de encuesta por cuestionario simple o entrevista.

Indicadores de esquizofrenia serían la sujeción a tratamiento médico, el internamiento y los síntomas de perturbación mental. Las técnicas adecuadas, encuesta, entrevistas clínicas y datos estadísticos y documentales de los centros de internamiento.

3) Menos de un 20 por 100 de los estudiantes que empiezan la carrera en España la acaban.

RESPUESTA

En esta hipótesis la única variable es "acabar la carrera de Medicina los estudiantes que empiezan". Como es una variable concreta o más bien un indicador, no se puede dividir en dimensiones e indicadores. El método de investigación sería fundamentalmente estadístico, mediante el examen de los expedientes de los alumnos matriculados en la Facultad de Medicina.

4) Los alumnos aventajados en ciencias no lo son en humanidades.

RESPUESTA

En este caso, para medir el aventajamiento en ciencias o humanidades, se podrían utilizar tests, además del examen de los expendientes de los estudiantes que abarcara la investigación.

5) Las mujeres son más conservadoras políticamente que los hombres.

RESPUESTAS

Aquí se trata de una variable intermedia en cuanto representa una de las dimensiones del conservadurismo. Indicadores podrían ser: lectura de revistas y periódicos conservadores. Asistencia a actos políticos de esta tendencia. Pertenencia a partidos de este tipo. Participación en actividades políticas conservadoras. Actitudes y opiniones, en fin, de este tipo. El método sería la encuesta por cuestionario simple o entrevista, y en cuanto a las opiniones y actitudes, la aplicación de escalas para medir el conservadurismo.

6) Las condiciones familiares influyen de manera importante en la delincuencia juvenil.

RESPUESTA

Entre las dimensiones de la primera variable están el ajuste conyugal, las condiciones económicas, las condiciones religiosas y morales. Indicadores, en general, serían, la separación, el divorcio, la falta de uno de los padres, el abandono de la fami-

lia, la embriaguez, las distensiones entre marido y mujer, el paro del cabeza de familia, el analfabetismo de los padres, la insuficiencia de ingresos. Dimensiones de la delincuencia juvenil podrían ser los distintos tipos de conducta desviada. Las técnicas apropiadas serían encuesta y estadística.

35 (A)

Para medir los efectos de la música de fondo sobre la retención del material verbal, se encomendó aleatoriamente a sujetos voluntarios (sacados de una clase de psicología introductoria sobre la base de su interés) el estudio de una corta historia en una de las tres condiciones siguientes:

a) Escuchando un disco X de música rock;

b) Escuchando un disco de música clásica con la segunda sinfonía de Beethoven, y

c) En un ambiente de silencio.

Todos los sujetos estuvieron al abrigo de ruidos extraños y a todos se aplicó el mismo test de elección múltiple sobre el recuerdo de detalles de la historia. El grupo que había estudiado en silencio recordó más detalles que cualquiera de los otros dos grupos expuestos a la música, y de ello se concluyó que la música es una fuente de distracción que conduce a una disminución de la retención.

A) Indicar la técnica de observación empleada en esta investigación.

RESPUESTA

Se puede considerar un experimento, en cuanto se somete a los tres grupos a una variable distinta, música rock, música clásica y silencio, y se quiere comprobar su efecto diferente sobre la retención de la historia.

B) Señalar cuál es la hipótesis, si existe, y, en caso afirmativo cuáles son las variables independiente y la dependiente.

RESPUESTA

La hipótesis expresada claramente en el enunciado, aunque no aparezca formulada de modo formal, es que la música de fondo influye en la retención del material verbal.

La variable independiente es la música de fondo, con tres categorías nominales: música de rock, música clásica y ausencia de música, si bien también se podrían considerar como tres variables independientes distintas.

La variable dependiente es la retención del material verbal.

C) Indicar cuáles son los hallazgos o resultados del experimento.

RESPUESTA

Los hallazgos de la investigación o los resultados del experimento se concretan en que el grupo que aprendió la historia en silencio recordó más detalles de la mis-

ma que cualquiera de los otros dos grupos expuestos a la música, de donde se infiere que la música es una fuente de distracción que disminuye la retención del material verbal.

D) Indicar las variables extrañas que hayan podido intervenir en el desarrollo del experimento y si han sido o no controladas.

RESPUESTA

Según el enunciado se halla controlada la variable ambiental de los ruidos extraños a la música de fondo, en cuanto se señala que todos los grupos estuvieron al abrigo de los mismos, si bien no se indica el procedimiento de control de esta variable utilizado, lo que no permite juzgar su validez.

Respecto a variables extrañas no controladas, del enunciado se deriva la existencia de dos, inherentes, la primera, al procedimiento de elección de la muestra, y la segunda, a la definición operativa adoptada de música. Nada se dice de otras posibles variables extrañas ni de su control, tales como las ambientales, las parásitas y las características de la demanda, las expectativas de los investigadores y la sospecha de su intento por los investigados. Es indudable que en este caso no se puede descartar, en principio, la posibilidad de que los investigadores comunicaran sutilmente sus expectativas, a los investigados, ni que éstos se formaran, sobre todo siendo voluntarios predispuestos a colaborar, una idea de los objetivos de la investigación y de la hipótesis que se pretendía contrastar, o que descubrieran de algún modo lo que se pretendía de ellos, todo lo cual pudo implicar la influencia de dichas variables extrañas.

En relación a la formación de la muestra investigada con sujetos voluntarios, supone un procedimiento de elección de la muestra sesgado y la obtención de una muestra sin seguridad estadística de que sea representativa. Hay que considerar, pues, si esta circunstancia de elección con voluntarios de la muestra puede dar lugar a distorsión en las respuestas al estímulo o variable independiente. No está clara esta influencia, es decir, el hecho de que por ser voluntarios pueda existir diferencia en la retención del material verbal según haya o no música de fondo, aunque tampoco se pueda descartar del todo.

En cuanto a las definiciones operativas, la música rock se hace representar por el disco X, la música clásica por el disco con la segunda sinfonía de Beethoven y la no música, por el silencio, y a partir de ello se concluye que la música distrae y disminuye la retención. Sin embargo, hay que tener en cuenta que no sólo los discos, sino los tipos de música rock y clásica son muy diversos y que no parece aceptable considerar simplemente ambos discos como representativos de la música rock y clásica. De la investigación se sigue que dichos discos impiden la retención del material verbal, pero no prueba que con todo tipo de música rock o clásica pase lo mismo.

Además, en la música se puede distinguir un componente de armonía y otro de sonido o ruido. Se confunden en ella, pues, dos factores y de la investigación no se puede saber si el impedimento observado se deriva de uno o de otro. Para controlar esta variable extraña habría sido necesario un nuevo grupo con ruido de los mismos decibelios que la música.

35 (B)

Una organización de publicación de revista de actualidad encargó a un investigador comprobar si existía alguna relación entre la lectura de revistas de actualidad y el estar bien informado sobre los acontecimientos del día.

A) Indicar la técnica de observación empleada. El investigador eligió cuidadosamente una muestra representativa de 1000 adultos, y obtuvo datos de ellos sobre el número de revistas de actualidad leídas en el mes último. Cada persona de la muestra fue sometida también a un corto test sobre acontecimientos actuales. Se halló una relación significativa: cuanto más revistas de actualidad leía una persona, tanto más alta era su puntuación en el test de sucesos actuales. Se concluyó que la lectura de revistas de actualidad hace estar a una persona bien informada.

RESPUESTA

Indudablemente se trata de una encuesta, mediante la cual se pretende averiguar si existe correlación entre la lectura de revistas de actualidad y el estar bien informado.

B) Señalar cuál es la hipótesis, si existe y, en caso afirmativo, cuáles son la variable o variables independientes y la dependiente.

RESPUESTA

La hipótesis que implica el enunciado, aunque no aparezca expresada en él de modo formal, es la existencia de alguna relación entre la lectura de revistas de actualidad y el estar bien informado. Las variables son dos: la lectura de revistas y el estar bien informado. La hipótesis establece simplemente una relación de variación conjunta entre ambas variables y sin indicar si se influyen entre sí. Por tanto de la hipótesis según está formulada no se puede determinar cuál sea la variable independiente y cuál la dependiente.

C) Indicar cuáles son los hallazgos o los resultados de la investigación.

RESPUESTA

El hallazgo de la investigación consistió en la relación entre la lectura de revistas de actualidad y la puntuación alta en el test sobre sucesos actuales, de donde se saca la conclusión de que la lectura de revistas de actualidad hace a una persona estar bien informada.

D) Indicar las variables extrañas controladas y no controladas en la investigación y dar explicaciones alternativas con base a estas últimas.

RESPUESTA

Del enunciado puede deducirse que se hallan controladas las variables extrañas ligadas a una elección de la muestra sesgada, ya que en él se dice que el investigador eligió *cuidadosamente* una muestra de 1.000 individuos, lo que hace suponer que por su tamaño y procedimiento de obtención no presentaba distorsiones.

En cuanto a las definiciones operativas de las variables, como la hipótesis sólo establece su correlación la variable lectura de revistas de actualidad, que en sí es un indicador definido operativamente, no representa más que a sí misma y por tanto, es en todo caso una definición operativa válida, si bien se puede hacer la objeción de que no se precisa qué se entendió por revistas de actualidad.

En cuanto a la variable estar bien informado de los acontecimientos actuales, el indicador utilizado para su observación, el test referente a los conocimientos de la actualidad, supuesto que aunque corto, sea completo y tenga validez y seguridad suficiente, es un indicador no contaminado en principio por variables extrañas que representa una definición operativa válida de la variable investigada.

Respecto a variables extrañas no controladas, no se puede descartar, en primer lugar, la actuación de las variables extrañas llamadas de escenario, y especialmente de la deseabilidad social y la aprensión evaluativa, que pueden haber inducido a muchos de los encuestados a exagerar el número y la intensidad de su lectura de revistas de actualidad. Sobre todo la gente más educada y por ello mejor informada puede haber pensado que quedaba en mal lugar si decía que no leía dichas revistas. Basándose, por tanto, en esta posibilidad, se puede dar la explicación alternativa de que los resultados son debidos a que gente bien informada quiere aparentar que lee revistas y no a una relación de dicha lectura con la buena información.

Según el enunciado, de los resultados obtenidos se infiere la conclusión de que la lectura de revistas de actualidad hace a una persona estar bien informada. Esta conclusión amplía indebidamente el hallazgo descubierto de la relación entre la lectura de revistas de actualidad declarada y el estar bien informado. No es admisible esta ampliación porque como suele suceder en los estudios correlacionales, existen otras muchas variables extrañas a la investigación no controladas, que pueden explicar también la mejor información. Entre ellas se pueden señalar, por ejemplo, que las personas que declaran leer más revistas de actualidad son también por ejemplo, las de un status económico más elevado, con mejores relaciones sociales, o con una educación más elevada, y son estas circunstancias y no la lectura de dichas revistas lo que les hace estar bien informados o que lean más revistas. Por otra parte, la mejor información en ellas puede proceder principalmente de otras fuentes, por ejemplo, lectura de la prensa diaria y no de dichas revistas, contactos personales, pertenencia a clubs, etc. (Estos ejercicios están basados en Paul C. Stern, 1979.)

8. LA EVALUACION DE LA INVESTIGACION

INTRODUCCION

En el capítulo anterior se ha examinado el diseño y dentro de él, de manera abstracta y general, los factores que pueden condicionar la validez del trabajo científico o la correspondencia de los resultados obtenidos mediante la observación con la realidad observada.

Corresponde, pues, tratar aquí ahora de la verificación de dicha validez. Sin embargo, hay que tener en cuenta que la palabra verificación ha adquirido en el campo científico, como se ha visto al estudiar el proceso de verificación, una significación concreta que agrupa el conjunto de operaciones de una investigación, mediante las cuales se contrasta con la realidad una hipótesis, ley o teoría. Por esta razón, en lugar de emplear aquí el término de verificación, se utilizará el de evaluación.

La evaluación se puede referir a las investigaciones de otros científicos o a las nuestras. En el primer caso se trata de la evaluación propiamente dicha y en el segundo, aunque también puede tener lugar una evaluación general como en las investigaciones ajenas, lo que se pretende más comúnmente es la comprobación de las distintas operaciones y de los resultados de la investigación realizada.

Entre los objetivos de esta disciplina, se deben incluir no sólo el capacitar para realizar adecuadamente una investigación en todas sus fases de planteamiento, trabajo de campo, clasificación, análisis e interpretación, sino también el facultar para enjuiciar y evaluar críticamente las investigaciones y estudios realizadas por otros profesionales y las nuestras propias. De ahí que, a continuación, se trate separadamente de la evaluación general de trabajos ajenos o propios y de la comprobación de las operaciones efectuadas en las investigaciones propias.

EVALUACION DEL TRABAJO CIENTIFICO

Evaluar se deriva etimológicamente, según el Diccionario de la Real Academia Española de *e* y *valuar*, señalar o calcular el valor de algo.

Como definición del término evaluación del trabajo científico puede servir su significación etimológica indicada y común: apreciar o estimar el valor de algo o, si se quiere, su validez, en este caso concreto de una investigación científica.

Por tanto, en la evaluación de un trabajo de investigación científica se deben distinguir dos aspectos fundamentales.

a) La evaluación de la corrección de las operaciones que comprende la investigación; y

b) La evaluación de sus conclusiones.

En cuanto al primer aspecto, parece obvio que si se trata de una investigación que pretende ser científica, se debe caracterizar por la aplicación en ella del método científico.

Por tanto, es en el método científico, es decir, en las condiciones y requisitos de las operaciones básicas que comprende, donde habrá que buscar los criterios para evaluar un trabajo de investigación determinado.

De acuerdo con el esquema de las operaciones básicas del método científico, puntos claves en los que se puede centrar la evaluación son los siguientes:

Determinación del problema de investigación. — Con relación a esta determinación, la evaluación deberá establecer qué clase de problema es y su significatividad teórica y científica y juzgar si está formulado correctamente, así como si es preciso en sus términos, si sus objetivos se hallan claramente definidos y si reune las condiciones exigibles.

Hipótesis. — En primer lugar se debe averiguar si existen hipótesis explícitas y cuáles son y si poseen las condiciones debidas. Si no es así se verá si las hipótesis se derivan del título de la investigación o de la exposición de la razón de ser y de los fines de la investigación.

Variables. — Halladas las hipótesis, el análisis de las variables que comprenden nos servirá para precisar las variables generales de la investigación y si el tipo de relación que une dichas variables es causal o no. A continuación, es preciso identificar los indicadores o variables empíricas observadas concretamente en representación de las variables genera-

les, así como buscar y examinar su definición operativa, datos que frecuentemente se pueden encontrar en la parte del informe referente al método empleado.

De modo especial se examinará si los indicadores y sus definiciones operativas representan adecuadamente a las variables generales objeto de investigación. En el caso de que no sea así, porque representen también en parte a otras variables extrañas a la investigación, se ofrecerán explicaciones alternativas basadas en dichas variables extrañas.

Diseño de la investigación.— Aquí se trata de determinar el tipo de diseño empleado para la práctica de la prueba implicada por la investigación, de su representación gráfica y de discutir su adecuación a la prueba de que se trate y las ventajas e inconvenientes que represente en relación con la misma.

Delimitación del campo de la investigación.— Se examinará críticamente respecto a este punto, los límites espaciales y temporales de la investigación, así como la definición de la población investigada y de modo especial si coinciden el universo real que se quiere investigar y el universo de trabajo efectivamente investigado.

Muestra.— Se identificarán, en su caso, la muestra utilizada, su representatividad y adecuación respecto a la población de la que se ha obtenido y los posibles sesgos derivados del procedimiento de elección empleado y del error derivado de su tamaño.

Si se detecta algún tipo de error sistemático en la elección de la muestra, que pueda viciar la representatividad de los resultados respecto a la población, se deberán imaginar las explicaciones alternativas consiguientes.

Instrumento de observación. — Se trata de examinar su corrección formal y de contenido, sobre todo en el caso de que haya sido construido específicamente para la investigación, y las garantías de validez y seguridad que ofrece.

Técnica de observación.— En relación a este punto se examinará la técnica de recogida de datos empleada, y sus características en este caso concreto y se juzgará su adecuación al problema investigado, así como las posibles incidencias que hayan tenido lugar en su aplicación y cómo se han solucionado.

Control de variables extrañas.— Se entiende por variables extrañas, los factores que aunque no son objeto de investigación, sin embargo,

pueden influir en las variables dependientes investigadas y, por tanto, en sus resultados.

Se examinará en relación a ellas de modo especial, si se han tenido en cuenta las que hayan podido mezclar sus efectos con los de las variables independientes objeto de la investigación y qué medidas se han adoptado para su control o neutralización de su influjo.

Análisis.— Aquí se trata de evaluar la pertinencia de los análisis estadísticos efectuados, la correcta ejecución de los mismos y la significatividad de los test estadísticos aplicados en su caso.

Variables extrañas no controladas y explicaciones alternativas.— Es posible que en una investigación científica no se hayan tenido en cuenta ni por tanto controlado variables extrañas que como tales pueden haber influido en los resultados de la investigación. Por tanto, en este momento se deberá identificar su existencia y basándose en cada una de ellas dar una explicación alternativa a los resultados obtenidos.

Evaluación de las conclusiones.— El análisis de los resultados debe servir de base a las conclusiones de las investigaciones, en cuanto éstas interpretan y expresan la significación y el alcance teórico de dichos resultados.

A este respecto la evaluación se debe centrar ante todo en los siguientes puntos:

1. Examinar si existen errores de interpretación y defectos lógicos en las consecuencias teóricas deducidas de los resultados obtenidos;
2. Juzgar si la posible generalización que supongan las conclusiones a otras situaciones o poblaciones no investigados, tiene o no justificación suficiente en los datos y resultados obtenidos en la investigación o si los extienden sin fundamento empírico suficiente, a otras situaciones o poblaciones no investigados.

La evaluación puede terminar con un juicio general final que sintetice y resuma la evaluación efectuada.

COMPROBACION

En toda investigación pueden aparecer errores. Del mismo modo que toda acción humana puede ser defectuosa, igualmente toda actividad investigadora puede presentar sus errores peculiares. Sin embargo, mientras se están realizando las operaciones de la investigación es difícil dar-

se cuenta de los que ocurren, con qué frecuencia se presentan y cuáles son, en conjunto y separadamente, sus efectos.

Por ello, la verificación de estos extremos requiere la revisión detenida, cuanto antes mejor, de las operaciones de cada fase de la investigación una vez terminadas, y particularmente actuaciones especiales respecto a los resultados generales.

En cuanto a estos últimos, las técnicas de comprobación se dividen en dos grandes categorías. La primera está formada, según Zarkovitch, por varios tipos de técnicas *a posteriori*, y la segunda por el uso de los métodos de muestreo. Además, los resultados deben ser objeto también de una comprobación teórica.

Dentro de las técnicas de verificación "a posteriori" se pueden destacar las siguientes:

a) Comparaciones con datos procedentes de fuentes independientes.

b) Estudios de consistencia.

a) *Comparación con datos procedentes de fuentes independientes.* Un procedimiento simple de comprobación de los datos recogidos en una encuesta, es comparar estos datos con los resultados disponibles procedentes de alguna fuente de información independiente o diferente de la encuesta en cuestión.

Esta comparación exige que los resultados que se utilizan como contraste sean homogéneos con los obtenidos en la encuesta por referirse a la misma población, época, estar originados por los mismos componentes, etc.

b) *Estudios de consistencia.* Pueden ser externos e internos. Los estudios de consistencia externa tienen por objeto ver cómo los datos recogidos en una encuesta concuerdan con algún conocimiento aceptado generalmente.

La verificación de la calidad de los datos por medio de un estudio de consistencia interna, implica una investigación del grado en que las estimaciones de las diferentes variables describen el fenómeno de la misma forma, o del grado en que concuerdan con otras variables con las que, deben ser congruentes matemáticamente o lógicamente.

En cuanto al uso en la comprobación de métodos de muestreo consiste en la práctica de una nueva encuesta a la misma población o parte de ella, en un tiempo que no esté alejado del de la realizada inicialmente.

Esta técnica de comprobación presenta las siguientes ventajas principales:

a) Su aplicación no depende de la existencia, a veces rara, de investigaciones y datos anteriores.

b) Es posible obtener mediante ella a veces un conocimiento de dónde aparecen los errores, bajo qué circunstancias, con qué características, lo que es muy valioso para planificar la mejora de las técnicas de recogida de datos; y

c) Las conclusiones sacadas de este tipo de verificaciones están basadas en las diferencias entre los datos recogidos en la encuesta original y los obtenidos en la encuesta de verificación. Esto permite llegar a estimaciones numéricas y hace posible su análisis matemático.

Sin embargo, esta técnica de verificación presenta también graves inconvenientes. Entre ellos destacan:

a) El hecho de que las circunstancias reales de la encuesta cambian de un momento a otro; y

b) El posible condicionamiento de las nuevas respuestas de los encuestados en la encuesta de verificación, por factores aleatorios y ocasionales de forma distinta que en la observación inicial.

COMPROBACION DEL PROCESO DE TEORIZACION

Respecto a la comprobación del proceso de teorización M. Bunge señala que toda teoría científica de alto nivel está sometida a cuatro baterías de prueba: empírica, interteórica, metateórica y filosófica. La primera es la que supone la investigación.

En cuanto a las pruebas interteóricas consisten en el examen de la compatibilidad de la teoría con el resto del saber científico a fin de asegurarse de su coherencia global.

El tercer examen, el de naturaleza metateórica, hace referencia a diversos caracteres formales, tales como la ausencia de contradicción, y semánticos, por ejemplo, la posibilidad de una interpretación en términos empíricos.

Finalmente, en cuanto a la cuarta prueba habrá que sospechar en particular de toda teoría que no esté de acuerdo con la metafísica dominante en círculos científicos.

EJERCICIOS

36 (A)

El Director de la Organización YMCA quería averiguar si el programa ayudaba a formar líderes. Revisó los registros de la organización y halló los nombres de 52 jóvenes de 10-12 años de edad, que habían sido miembros de la organización en 1953. Utilizando los registros escolares de dicho año y de la misma localidad, Hometown, eligió al azar 52 muchachos de la misma edad, excluyendo de esta lista los muchachos comprendidos ya en la lista Y. Intentó contactar con los 104 individuos, ya hombres en la fecha de la investigación: 1980, y pudo localizar sólo a 48 miembros de Y y 41 de no miembros y determinó su ocupación presente. Empleando un índice estándar de la categoría socioeconómica de las ocupaciones, estableció una puntuación para cada sujeto de su posición socioeconómica, de acuerdo con su ocupación presente. Como el grupo Y alcanzó el promedio de puntuación más alto, concluyó que el programa Y formaba líderes.

Se pide evaluar esta investigación y para ello,

a) Discutir la corrección del problema a investigar en su formulación.

RESPUESTA

Está formulado claramente y es preciso en sus objetivos. Reúne también las condiciones de ser concreto y verificable, no plantear juicios de valor y presentar una cierta trascendencia y novedad.

b) Señalar cuál es la hipótesis, si existe, y si reune las condiciones exigibles.

RESPUESTA

La hipótesis se deriva claramente del enunciado e implica la suposición de que la participación en el programa Y incrementaba la probabilidad de que un joven llegara a ser líder de mayor. Se trata de una hipótesis científica –su contraste requiere una investigación–, de dos variables en relación de dependencia y por tanto causal, simple con un solo enunciado relacional, y de trabajo.

Reúne de algún modo las condiciones de claridad, comprensión, sus términos poseen una realidad empírica susceptible de verificación, puede tener relación con la teoría del liderazgo ya existente y tiene un cierto alcance general.

c) Indicar las variables que forman la hipótesis y la definición operativa de las mismas adoptada.

RESPUESTA

Variable independiente: Participación en Y. Definición operativa: ser un muchacho incluido en la lista de participación en Y de 1953, con una edad de 10-12 años.

Variable dependiente: Liderazgo. Definición operativa: La puntuación de los sujetos, según su ocupación actual, en un índice estándar de la categoría socioeconómica de las ocupaciones.

d) Especificar el tipo de diseño que supone la investigación y representarlo gráficamente.

RESPUESTA

Existen dos grupos, uno que sufre el impacto de la variable independiente y el otro no. Ambos grupos, no igualados, son objeto de una sola medición y formados mucho tiempo después de haber ejercido su influjo la variable independiente. Se trata, por tanto, de un diseño *ex-post-facto*, que se podría representar como sigue: (Los puntos suspensivos indican la formación y medición de los grupos mucho tiempo después del estímulo.)

$$X \ldots 0_1$$

$$\ldots 0_2$$

e) Examinar la delimitación del campo de la investigación.

RESPUESTA

La investigación se puede considerar delimitada con precisión en el tiempo, un año concreto y determinado; y en el espacio, una localidad de los Estados Unidos: Hometown. No se alude expresamente en el enunciado a la definición de la población investigada, pero se puede inferior que ésta se extiende a todos los muchachos que en el año de 1953 tenían de 10 a 12 años de edad.

f) Enjuiciar el instrumento de observación utilizado.

RESPUESTA

Este instrumento no ha sido construido para esta investigación, sino que se utiliza uno ya existente: el índice estándar que se indica en el enunciado. Es de suponer que este instrumento al tratarse de un índice estándar, aunque nada se dice al respecto, sea suficientemente válido y seguro, y que esta validez esté además contrastada por su aplicación satisfactoria en muchos otros casos.

g) Indicar si se ha utilizado muestra en la investigación y, en su caso, enjuiciarla críticamente.

RESPUESTA

Si existen muestras formadas por 104 hombres: 52 que fueron miembros de YMCA y 52 de la misma edad y asistencia a escuelas públicas, pero no miembros de Y.

Población: Personas de Hometown que en 1953 tenían de 10-12 años de edad.

Sesgo de la muestra: La muestra de jóvenes de las escuelas públicas obtenida no es representativa de todos los jóvenes de dicha edad de la ciudad, porque podía haber una proporción más o menos elevada de jóvenes que estudiaron en dicho año en otro tipo de escuelas y porque no sabemos si los YMCA eran todos también de escuelas públicas. Resulta así que este muestreo confunde la pertenencia a Y con el tipo de escuela a la que se asiste.

Otra fuente de sesgo consiste en que no todos los sujetos de la muestra aparecen en los resultados. Hay 4 de los Y y 9 de los restantes no localizados y no sabemos su puntuación en liderazgo.

Esta circunstancia da pie para la explicación alternativa de que la ciudad estaba deprimida y que los jóvenes no localizados, con altas cualidades y ambiciones, emigraron a otros lugares. La atrición o pérdida de elementos de la muestra puede esconder una variable extraña.

Por último, en el enunciado se generalizan los resultados y se sostiene que el programa Y, en general, y por tanto sin distinción de poblaciones, produce líderes. Esta generalización no está fundada, aunque la muestra no fuera sesgada. La conclusión va más allá de la población de los jóvenes de Hometown, que en 1953 tenían 10-12 años.

h) Indicar y enjuiciar los hallazgos o resultados de la investigación.

RESPUESTA

Como en el enunciado no se exponen los análisis realizados, la evaluación a este respecto se tiene que limitar a enjuiciar los resultados que se dicen obtenidos.

Los resultados en cuestión revelan concretamente que los que fueron miembros de YMCA lograron un estatus más elevado en su ocupación que los no miembros, asistentes a escuelas públicas, de la misma edad y sexo y en la misma localidad de Hometown. De este resultado se infiere la conclusión de que el programa YMCA favorece la formación de líderes.

i) Identificar las variables extrañas controladas o no que hayan podido actuar en la investigación y expresar las explicaciones alternativas que pueden basarse en ellas.

RESPUESTA

Variables extrañas sujetas a algún control. Entre ellas las siguientes fueron mantenidas constantes, o iguales para todos los investigados; la edad de los sujetos cuando se mide su estatus; el sexo y la ciudad en que vivían cuando tenían 10-12 años.

Sin embargo, según el enunciado, ninguna variable se controló mediante aleatorización. Los sujetos no fueron asignados al azar para participar o no en el programa Y. Esta circunstancia es causa de que existan en la investigación, como veremos, un número de factores que pudieron variar de forma incontrolada.

Variables externas no controladas y explicaciones alternativas. Se advierte, en primer lugar, una definición inválida del liderazgo. Aunque es verdad que se puede lograr a veces en la ocupación un alto nivel por las cualidades personales de liderazgo, también se puede alcanzar esta posición a través de conexiones sociales, una preparación especial posterior o una personalidad agradable. Estos factores son todos variables extrañas que pueden explicar la puntuación alta de los sujetos en el índice utilizado. Este, por tanto, no mide específicamente el liderazgo. En realidad, su medida es difícil. Exigiría, por ejemplo, el diseño de una situación típica en la que se pudiera comprobar experimentalmente la capacidad de los investigados como líderes.

Variables orgánicas. Se trata en el ejercicio de un caso clásico de correlaciones que se usan para probar causas. No es seguro que el liderazgo fuera causado por el

programa Y porque existe un número de variables orgánicas, correlacionadas posiblemente con la pertenencia a la Organización YMCA que pueden ser más importantes. He aquí algunas que sugieren explicaciones alternativas de los resultados.

El estatus socioeconómico de los padres: si YMCA era en 1953 una organización utilizada principalmente por las clases media y alta de Hometown, los jóvenes de YMCA tuvieron una ventaja sobre los de las escuelas, incluso antes de entrar en Y. Su éxito posterior pudo ser debido, por ejemplo, a la influencia de su familia en la ciudad o a su socialización en unos patrones de conducta favorables al éxito social.

Igualmente pudo ocurrir que YMCA estuviera situada en una parte residencial de la ciudad y, por ello, recibió una muestra seleccionada de jóvenes que vivían cerca, aun sin excluir las clases inferiores.

Por otra parte, los muchachos del grupo de comparación asistieron todos a las escuelas públicas, mientras que no sabemos si esto es verdad en los del grupo YMCA, que pudieron ser en gran parte alumnos de colegios privados y parroquiales, lo que supone una desventaja para los primeros.

Son plausibles, por tanto, otras explicaciones alternativas distintas a la inferida en la investigación.

Así se puede sostener que pertenecieron a YMCA niños que procedían de familias con un estatus alto y alcanzaron un nivel en la ocupación más elevado por conexiones familiares, mayor posibilidad de acceso a la educación y a la cultura y por otras condiciones personales y sociales y no sólo por su formación en liderazgo.

Un diseño experimental que hubiera asignado aleatoriamente unos jóvenes al programa Y y otros al grupo de control, podría haber sido una forma de control de estas variables orgánicas, pero aquí fue, en todo caso, impracticable, al tratarse de una investigación *ex-post-facto*. No obstante, en la formación de los grupos se podía haber procurado que fueran iguales en algunas de las variables importantes: estatus de los padres, asistencia a la escuela pública, etc.

Por último, hay que señalar que dado el largo período de tiempo que media entre la participación o no en el programa YMCA y la medición del nivel en ocupación, hay que tener en cuenta el factor tiempo y las variables ligadas al mismo, especialmente la historia y la maduración. Los acontecimientos familiares y personales y la evolución física y psíquica de los miembros de ambos grupos puede haber sido en dicho período muy diversa y suficiente para explicar en todo o en parte las diferencias entre los grupos detectadas en la investigación.

36 (B)

Un investigador supuso que los hombres rinden más en condiciones de competición que las mujeres. Para contrastar este supuesto, se encargó a un grupo de estudiantes universitarios, pertenecientes a un curso determinado, la resolución de una serie de problemas matemáticos, con la indicación de que formaban parte de un test de inteligencia y que sus puntuaciones serían comparadas con las de otros estudiantes de la misma universidad (se estimó que estas instrucciones tenían como fin hacer que la situación fuera competitiva). Los hombres en este grupo cumplieron significativamente mejor que las mujeres.

A un segundo grupo de estudiantes, pertenecientes al mismo curso, se encomendó el aprendizaje de memoria de una serie de sílabas sin sentido, con instrucciones de que el estudio se refería a la relativa dificultad de aprender diferentes clases de material sin sentido (estas instrucciones se consideraron no competitivas). En este

grupo, los chicos y las chicas cumplieron igualmente bien. El investigador concluyó que los hombres respondían más eficazmente que las mujeres en situaciones competitivas.

Se pide evaluar estas investigaciones y para ello:

a) Discutir la corrección del problema a investigar en su formulación.

RESPUESTA

La formulación y los objetivos son claros, y el problema es concreto y verificable pues no plantea juicios de valor y presenta una cierta novedad, si bien se puede advertir una falta de precisión en la forma de su desarrollo.

b) Señalar la hipótesis si existe y si reúne las condiciones exigibles.

RESPUESTA

Existe una hipótesis expresa que consiste en la suposición de que los hombres rinden más en condiciones de competición que las mujeres. Se trata de una hipótesis científica de trabajo, pues su contraste se pretende realizar mediante la investigación proyectada, y de dos variables en relación de dependencia y, en este sentido, de carácter causal.

Reúne las condiciones de claridad y comprensión, sus términos poseen una realidad empírica susceptible de verificación, puede tener una trascendencia teórica y presenta un alcance general.

c) Indicar las variables que forman la hipótesis y la definición operativa de las mismas adoptada.

RESPUESTA

Una variable independiente es el sexo, que es una variable empírica que no necesita definición operativa que la especifique. Otra variable independiente es la competición, con dos categorías: la situación competitiva y la no competitiva. La definición operativa de la competición se concreta en el caso de ambas categorías en las distintas instrucciones dadas por los investigadores para la realización de los problemas y trabajos propuestos. La variable dependiente son los resultados de las pruebas realizadas.

d) Examinar la delimitación del campo de la investigación.

RESPUESTA

No está precisado este extremo en el enunciado y, por tanto, no se puede juzgar el alcance geográfico ni temporal de la investigación. Tampoco aparece concretada la población investigada, pues nada se especificó del nombre de la Universidad, de la carrera universitaria ni del curso y su composición según el sexo.

e) Especificar el tipo de diseño que supone la investigación y representarlo gráficamente.

RESPUESTA

La investigación comprende dos grupos, respecto de los cuales en el enunciado no se dice que estén elegidos al azar. Ambos grupos son sometidos a un estímulo o variable experimental distinta y son objeto de una sola medición después del estímulo.

Se trata de un diseño de tipo experimental bastante imperfecto, que gráficamente hay que representar por:

$$X \quad O_1$$

$$X' \quad O_2$$

f) Enjuiciar el instrumento de observación utilizado.

RESPUESTA

Según el enunciado, únicamente se le puede considerar como prueba más bien académica genérica, pues nada se dice que se tratara de test específicos y determinados, y, por tanto, de validez y seguridad no conocida.

g) Indicar si se han utilizado muestras y criticar, en su caso, el procedimiento de su elección y el de formación de los grupos.

RESPUESTA

Del enunciado puede deducirse que existen dos muestras, en cuanto los grupos con los que se trabaja parece forman parte de un mismo curso universitario. Pero como se desprende del texto que los grupos se formaron por las buenas, hay que considerar su elección casual y no representativa.

h) Indicar y enjuiciar los hallazgos o resultados de la investigación.

RESPUESTA

Como en el enunciado no se exponen las tablas formadas y los análisis realizados, la evaluación a este respecto se tiene que limitar a indicar los resultados y las conclusiones que se derivan de ellos.

Resultados: los de los chicos en la situación competitiva fueron mejores que los de las chicas e iguales en la no competitiva.

La conclusión que se infiere es que los hombres respondían más eficazmente que las mujeres en situaciones competitivas, confirmándose la hipótesis.

i) Identificar las variables extrañas, controladas o no, que hayan podido actuar en la investigación y exponer las explicaciones alternativas que puedan basarse en ellas.

RESPUESTA

Variables externas controladas son únicamente, según el enunciado, la Universidad y el curso que es el mismo en los dos casos; por tanto, permanecieron constan-

tes a lo largo de la investigación, por lo que no se pudieron derivar diferencias en los resultados de los grupos debidos a ellas.

Nada se dice de que se controlara por aleatorización la composición de los individuos que comprendieron los grupos investigados.

Variables externas no controladas son, en principio, todas las demás, excepto las ligadas al transcurso del tiempo entre las diversas observaciones, pues, suponiendo que la observación de los dos grupos se hiciera simultáneamente, al existir una sola observación de cada grupo es imposible que actúen estas variables.

En especial, merecen subrayarse las variables externas siguientes.

La definición operativa de la variable independiente, competitividad, basada únicamente en las instrucciones dadas, resulta pobre y discutible en sus efectos como tal. Por otra parte, ambos grupos difieren no sólo en las instrucciones que recibieron, competitivas o no, sino también en las tareas que se les asignaron o en el tipo de material usado para comprobar sus realizaciones. De ahí que toda diferencia en el sexo atribuida a la competición puede ser asignada por iguales razones al tipo de tarea: los hombres, son mejores en la resolución de problemas matemáticos y las mujeres en tareas verbales.

Esta circunstancia se puede poner también en conexión con el hecho de que la variable sexo es una variable orgánica, con la cual pueden estar otras variables parásitas relacionadas diferencialmente. Si los hombres son mejores en matemáticas y las mujeres tienden a destacar en las tareas verbales, esta circunstancia es suficiente para explicar los resultados de la investigación, independientemente de la situación competitiva o no. Las dos tareas deberían haber sido iguales en dificultad y, en ambas, los hombres y las mujeres deberían haber sido iguales inicialmente; para comprobarlo debería haberse realizado algún pretest.

Por otra parte, no se sabe cómo fueron elegidos los dos grupos, sólo se dice que proceden de un curso universitario, pero sin indicar las bases de su procedencia. No se puede descartar por ello, que, por alguna razón, un grupo fuera más verbal, más matemático o más competitivo que el otro, según el sexo.

Nada se dice tampoco del control de variables extrañas tales como las características de la demanda y las expectativas de los investigadores. En este caso, no se puede descartar por principio la posibilidad de que los investigadores comunicaran sutilmente sus expectativas a los investigados, ni de que éstos se formaran una idea de los objetivos de la investigación y de las hipótesis que se pretendía contrastar, o que descubrieran de algún modo lo que se pretendía de ellos, todo lo cual pudo implicar la influencia de dichas variables extrañas.

Por último, según el enunciado, de los resultados obtenidos se infiere la conclusión de que los hombres respondan más eficazmente que las mujeres en condiciones de competición. Esta generalización es excesiva. Los resultados sólo establecen, a lo sumo, que en las situaciones a que se refiere la investigación los hombres en el grupo competitivo cumplieron mejor que las mujeres y nada más, pues esta conclusión no se puede extender a otras situaciones competitivas ni tampoco a otros grupos porque no consta que ni la situación ni el grupo investigado fueran representativos de todas las múltiples situaciones competitivas posibles y de otros grupos distintos. (Estos ejercicios están basados en Paul C. Stern, 1979.)

Sección 3

MUESTRAS

MUESTRAS

9. CONDICIONES Y ELEMENTOS DE LAS MUESTRAS

NOCION

Una vez definido el problema a investigar, formuladas las hipótesis, determinadas las dimensiones e indicadores de las variables y delimitado el campo de la investigación, se hace preciso realizar, además de las anteriores, otra opción antes de empezar con el trabajo de campo, la recogida de datos mediante la aplicación de los métodos elegidos.

Esta nueva opción consiste en la decisión de si las unidades objeto de observación o estudio van a ser todas las que forman el universo, o únicamente se va a extender la indagación a una parte representativa o muestra de aquéllas.

Salvo en el caso de poblaciones pequeñas, esta última solución es la que se impone en la realidad, dado que en los demás casos, razones de tiempo, coste y complejidad de las operaciones de recogida, clasificación y análisis de los datos, descartan en la práctica la posibilidad de que el estudio abarque individualmente a todas las unidades que comprende el universo.

Por ello, al constituir el uso de las muestras una técnica de aplicación casi general en las investigaciones sociales, es evidente la importancia que tiene su estudio y la necesidad en que se halla el investigador social de conocer por lo menos sus principios y prácticas básicos, aunque se trate de una materia, basada en las leyes de azar y el cálculo de probabilidades, que pertenece al campo matemático de la estadística.

Una muestra es simplemente, en general, una parte representativa de un conjunto, población o universo, cuyas características debe reproducir en pequeño lo más exactamente posible.

De modo más científico, se pueden definir las muestras como una parte de un conjunto o población debidamente elegida, que se somete a observación científica en representación del conjunto, con el propósito de obtener resultados válidos, también para el universo total investigado.

Las muestras tienen un fundamento matemático estadístico. Este consiste en que obtenidos de una muestra, elegida correctamente y en proporción adecuada, unos determinados resultados, se puede hacer la inferencia o generalización, fundada matemáticamente, de que dichos resultados son válidos para el universo del que se ha extraído la muestra, dentro de unos límites de error y probabilidad, que se pueden determinar estadísticamente en cada caso.

Las muestras presentan las siguientes ventajas, en las ciencias sociales:

1. Mediante ellas, con una muestra relativamente reducida con relación al universo, se pueden encuestar las grandes poblaciones y núcleos humanos, que de otra manera sería muy difícil o prácticamente imposible investigar.

2. En todo caso las muestras suponen una gran economía en las encuestas y la posibilidad de mayor rapidez en su ejecución.

3. Además, como señala Sheuch, en el "Tratado de Sociología empírica", de R. König (Tecnos), las muestras ofrecen otra ventaja, muy importante y casi inadvertida; una muestra puede ofrecer resultados más precisos que una encuesta total, aunque esté afectada del error que resulta de limitar el todo a una parte.

En una muestra se puede prestar más atención al caso particular. De ello resulta eventualmente una limitación de los errores fácticos de tal orden, que la muestra puede conducir a un resultado más preciso en su totalidad que una encuesta total. De ahí que se utilicen muestras en las estadísticas americanas para comprobar el grado de exactitud de las encuestas totales o censos.

CONDICIONES DE LAS MUESTRAS

Las condiciones fundamentales de las muestras son cuatro:

1. Que comprendan parte del universo y no la totalidad de éste.

2. Que su amplitud sea estadísticamente proporcionada a la magnitud del universo. Esta condición se halla en relación con el punto práctico de determinación del tamaño de la muestra y sirve para decidir si, según las unidades que comprende respecto al universo, una muestra es o no admisible.

3. La ausencia de distorsión en la elección de los elementos de la muestra. Si esta elección presenta alguna anomalía, la muestra resultará por este mismo hecho viciada.

4. Que sea representativa o reflejo fiel del universo, de tal modo que reproduzca sus características básicas en orden a la investigación. Esto quiero decir que si hay sectores diferenciados en la población que se supone ofrecen características especiales, a efectos de los objetivos de la investigación, la muestra también deberá comprenderlos y precisamente en la misma proporción, es decir, deberá estar estratificada como el universo.

Además de estas condiciones fundamentales, se ha de tener en cuenta que las muestras se basan, como han destacado algunos autores, en una teoría del público como masa atomizada (Bordieu), producen la ruptura de las redes topológicas y arrancan al individuo de la urdimbre social (Ibáñez) y prescinden de rasgos sociales estructurales importantes (Scheuch). Por ello, el empleo de las nuestras debería implicar una información previa sobre la estructura de la colectividad investigada.

Una cierta solución a este problema son los llamados sondeos contextuales, como, por ejemplo, el muestreo gradual, para conocer la opinión de los militantes de un partido, de, primero, federaciones provinciales, segundo, de secciones locales y, tercero, de militantes de éstas. Se ha de recordar que el individuo no está nunca totalmente desconectado de la sociedad, sino que, aunque actúe aisladamente de ésta, lleva en sí el influjo de los grupos sociales a los que pertenece. (Ver ejercicio 37)

BASE DE LA MUESTRA

Los elementos principales de la muestra, son la base y la unidad de la misma.

La base de la muestra es, en todo caso, la población de la que se obtiene o saca la muestra. Esta población puede no estar censada o puede estarlo. En el primer caso, existe la base de la muestra en sentido real o material, pero no en sentido formal, mientras que en el segundo caso hay base de la muestra tanto material, la población real a investigar, como formal, el censo de todos los individuos que la componen. Precisamente con el término de base de la muestra se suele designar de manera específica este censo o registro.

La importancia de la base de la muestra, tomada específicamente o en sentido estricto, se deriva de que la muestra, operativamente, es el resultado de la elección de unidades dentro de una población o conjunto previamente determinado de aquéllas. Por ello, fundamento básico de la muestra, es la existencia de un registro de dicho conjunto, en el que aparezcan individualizadas todas sus unidades, y permita realizar la elección mediante un sorteo riguroso.

Esta concreción individualizada de las unidades del universo, es el punto de partida y el fundamento necesario para realizar con rigor la elección en que consiste la muestra y por ello se dice que constituye la base de la muestra en sentido estricto.

Esta puede consistir en un censo, un registro, una lista, un fichero, un catálogo, un mapa, un plano, etc. Se debe dar normalmente un número a cada ficha o elemento del catálogo.

En la base de la muestra deben figurar individualizadas todas las unidades que forman la población con expresión de su número en el universo, nombre, domicilio en su caso, etc. Por ello, la base de la muestra hace posible la identificación de los elementos que se hayan seleccionado mediante la muestra y su encuesta posterior.

La base de la muestra en este sentido estricto, no siempre existe. Hay muchos universos que no están censados o catalogados y que es prácticamente imposible catalogar. Por ejemplo, no lo están el público que circula por las calles ni los asistentes a un espectáculo. La solución que se adopta entonces es practicar la elección de la muestra por algún procedimiento aleatorio imperfecto, como v.g. encuestar uno de cada cinco que se encuentren por la calle

Lo ideal sería formar una base de muestra específica para la encuesta. Pero esto no es posible, por lo general, en las investigaciones sociales. Supondría casi tanta dificultad como efectuar la encuesta a toda la población. Por ejemplo, si para hacer una encuesta que alcanzase a toda España se decidiese formar una base de la muestra propia, esto supondría realizar un censo de todos los españoles.

Por esta razón, normalmente se utilizan bases ya formadas. En este caso es preciso asegurarse de que la base está actualizada, pues si no nos exponemos a dejar fuera de la muestra, por ejemplo, a la población más joven de ella.

Además hay que verificar si la población que queremos encuestar (universo teórico) y aquella de la que se toma de hecho la muestra (universo de trabajo) o a la que se refiere la base de la muestra, coinciden, porque podría haber discrepancias más o menos extensas entre ellas.

En el orden práctico, respecto a la base de la muestra, hay que buscarla, analizarla y prepararla a efectos del muestreo si existe. Si no existe, y ello es posible, hay que formarla. En este caso se debe cuidar, de modo especial, según lo indicado, que no resulte ordenada de alguna forma que pueda determinar una elección distorsionada o no representativa de la muestra.

Por último hay que advertir que algunos autores, entre ellos Sánchez Crespo (1976, p. 6), llaman a la base de la muestra marco.

Además de este sentido estricto que he indicado, entiende también por marco, en sentido amplio, la información previa, mapas, comunicaciones, etc. que permita llegar a todos los elementos de la población a encuestar. (Ver ejercicios 38, 41)

UNIDAD DE LA MUESTRA

La unidad de la muestra es cada uno de los elementos que comprende su base y figuran, numerados e individualizados, en el registro de la misma cuando existe.

Las unidades de este registro representan y remiten a las unidades de observación que forman el universo a investigar.

Como he indicado la base formal, estricta, y sus unidades hacen posible la identificación de las unidades de observación del universo a encuestar.

Las unidades de la muestra no solo pueden ser simples, sino también colectivas, como cuando están constituidas por familias, grupos, sectores de ciudades, pueblos, etc. El número de individuos que comprende la unidad de la muestra cuando es colectiva se llama talla de la muestra.

Esta talla de las unidades de la base de la muestra o población es lo que determina lo que se suele llamar el nivel de agregación en la investigación. Este nivel puede presentar grados de agregación diversos, bien de carácter espacial, social, económico, según los colectivos que se elijan como agregados.

Así, por ejemplo, para estudiar las compras de TV en color se puede operar al nivel de agregación de individuos o de colectivos y, dentro de los colectivos, especialmente, estos pueden ser municipios, distritos, regiones, naciones. Desde un punto de vista social los agregados pueden ser familias, establecimientos de venta, empresas, organismos y asociaciones determinados.

RELACION ENTRE UNIVERSO Y MUESTRA

El diagrama que sigue de Mora y Araujo (1973, p. 12), muestra gráficamente la relación entre universo y muestra.

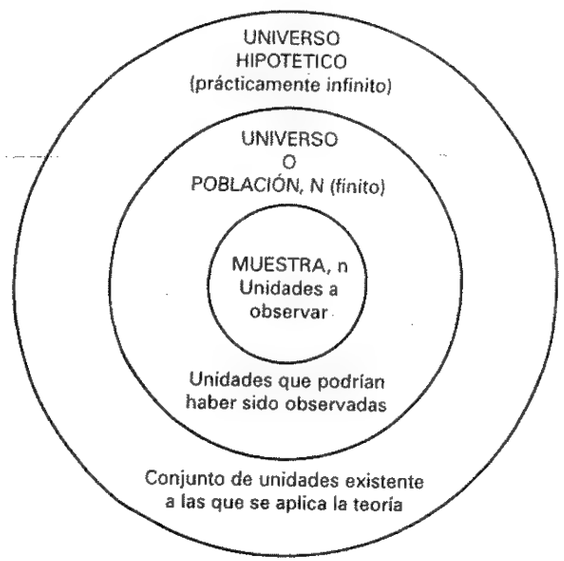
La muestra, como se ve, es una parte reducida de una determinada población o universo, en relación a la cual representa cuantitativamente una fracción.

Pero también la muestra tiene la misma forma circular que el universo. Ello pretende indicar que, cualitativamente, la muestra debe tener las mismas características que la población que representa.

El universo está formado por toda la población o conjunto de unidades que se quiere estudiar y que podrían ser observadas individualmente en el estudio.

Sin embargo, por razones de economía se investiga en su representación la muestra.

En cuanto a la relación entre universo y muestra, hay que tener en cuenta que, como subraya Sánchez Carrión (1995, 107) "todo lo relacionado con poblaciones no observadas tiene que ver con las *probabilidades*. Sólo de lo que se observa (la muestra) se puede hablar con certeza; de lo no observado (la población) siempre hay que hablar en términos probabilísticos: hemos preguntado a esta poca gente y opina esto; si hubiéramos preguntado a todos, *lo más probable* es que hubiéramos obtenido tal o cual resultado".



En el gráfico también se ve cómo el universo o población forma parte también de un conjunto más amplio al que se llama universo hipotético. Este universo hipotético está constituido por todas las poblaciones existentes en las que se podría también efectuar la investigación proyectada.

Precisamente lo que da validez máxima a un hallazgo científico es haber sido comprobado no sólo en una población sino en relación a muchas de las existentes.

Entre el universo y la muestra se puede distinguir una relación cualitativa y otra cuantitativa.

A la relación cuantitativa se refieren los conceptos de fracción de muestreo y de coeficiente de elevación. La fracción de muestreo es el porcentaje que representa la muestra respecto al universo.

Se obtiene dividiendo la muestra por el universo. Por ejemplo, si en un universo de 3.000 unidades hemos elegido una muestra de 600, la fracción de muestreo será 0,20 o el 20 % .

Esta fracción puede tener diversas aplicaciones entre ellas la de hallar la composición por estratos de una muestra cuando se conocen la composición por estratos del universo o población.

En cuanto al coeficiente de elevación no es otra cosa que el número inverso a la fracción de muestreo y consiste en la cantidad por la que hay que elevar la muestra para obtener el universo.

Se obtiene, al reves de la anterior, dividiendo el universo por la muestra. En el caso indicado sería 5. Prácticamente el coeficiente de elevación señala las veces que la muestra está contenida en el universo.

La relación cualitativa exige, como se ha indicado, que el universo y la muestra reunan las mismas características básicas en igual proporción. Ello es importante sobre todo en los universos estratificados, que comprenden categorías diversas.

Esta correspondencia es difícil que en la práctica se de exactamente, sobre todo porque en las encuestas siempre hay fallos: entrevistas previstas que no se pueden realizar o no son válidas, cuestionarios que no se devuelven, etc.

Por ello, para que se conozca la discrepancia existente, es importante formar un cuadro comparativo en el que consten los porcentajes que supone cada categoría en el universo y en la muestra. A esta operación se le llama verificar el equilibrio de la muestra.

Esta operación muestra a veces que los porcentajes de los estratos en las muestras, sobre todo una vez realizada la encuesta, difieren bastante de los del universo.

Hay casos en los que esta discrepancia se procura deliberadamente, a fin de que un estrato que interesa estudiar especialmente y se considera insuficientemente representado en la muestra lo esté más. Para ello se aumenta el porcentaje que le corresponde y se disminuye el de otros estratos en la misma proporción. En estos casos, se debe corregir esta discrepancia en los resultados, ponderando éstos en la muestra según los porcentajes que corresponden a cada estrato en el universo. (Ver ejercicios 39, 40, 43, 44)

EJERCICIOS

37

Dadas las investigaciones que se indican a continuación, se pide determinar razonadamente si se basan en una muestra del universo correspondiente y si, en caso afirmativo, la muestra es correcta.

A) A efectos de realizar el sociograma de una clase, se ha pedido a todos los alumnos que expresaran en una papeleta los nombres de sus compañeros de clase que les gustaría tener sentados a su lado y aquellos que no.

RESPUESTA

Es claro que no constituye una muestra, en cuanto el universo son todos los alumnos de la clase y todos ellos fueron encuestados.

B) Para estudiar las prácticas sexuales de los varones en una prisión se entrevistó a todos los que se presentaron voluntariamente a responder al cuestionario que se había preparado.

RESPUESTA

Genéricamente se puede decir que, en este caso, hay una muestra, en cuanto sólo se hizo la encuesta a una parte del total de internados en la prisión. Sin embargo, se trata de una muestra viciada, basada en un sistema de elección inadecuado, por lo que no se puede considerar representativa del universo, ni sus resultados extendidos a éste.

C) Un antropólogo social ha convivido durante dos años con una familia típica de una localidad de las Hurdes, se ha ganado su confianza y ha logrado que sus miembros le expusieran los aspectos de su vida de interés para su investigación.

RESPUESTA

Este estudio, aunque se diga que se basa en una familia típica, no se puede considerar como muestra, pues un solo caso no es suficiente al efecto.

D) Para estudiar las infracciones de circulación cometidas por no detención ante un "stop", un equipo de observación ha permanecido, de ocho de la mañana a ocho de la tarde, ante el "stop" durante tres días de la semana consecutivos, y los observadores consignaron los detalles de una por cada tres infracciones observadas.

RESPUESTA

El universo son todas las infracciones. Como sólo se investiga una parte, se puede hablar de muestra de todas ellas. Pero esta muestra es desviada y no representativas del universo, porque sólo proporciona información de unas horas determinadas y de sólo tres días consecutivos a la semana, sin que nada diga sobre las restantes horas del día y los demás días de la semana, en los que el número y calidad de las infracciones puede ser diferente.

E) Para predecir los resultados en una comunidad de una elección municipal que iba a tener lugar unos días después, el encuestador preguntó su candidato preferido a todos los hombres y mujeres con derecho a voto.

RESPUESTA

No existe muestra, pues se consulta a todo el universo, los hombres y mujeres con derecho a voto y no a parte de ellos. Pero, por otra parte, el universo no es representativo a los efectos de la encuesta, ya que se dice, se consulta indiscriminadamente a todas las personas con derecho a voto y hay que tener en cuenta que siempre suele haber un porcentaje a veces muy elevado que no votan. En consecuencia, sus preferencias viciarían la predicción, pues los resultados efectivos se basan únicamente en la conducta de los que realmente tengan intención de votar y voten.

F) En otra elección a diputados, en la que se presentaban los candidatos, uno de derechas y otro de izquierdas, se realizó el sondeo de opinión, a una muestra elegida al azar con base en la lista telefónica, por medio del teléfono. Se obtuvo un resultado favorable al candidato de derechas, aunque resultó luego elegido el de izquierdas.

RESPUESTA

En este sondeo la muestra tampoco es representativa, aunque se haya elegido al azar. Presenta la distorsión que supone el hecho de que los que poseen teléfono son gente de un cierto nivel económico, y por ello es de suponer que sean en mayor proporción de derechas. Ello explica la falta de acierto de la encuesta del resultado efectivo de las elecciones.

G) Para descubrir la existencia de "cliques" en una clase numerosa de la universidad, el investigador preguntó a una muestra elegida al azar de estudiantes el nombre de sus amigos más íntimos.

RESPUESTA

Si se basa en una muestra, según el enunciado, pero esta muestra es improcedente en relación al fin propuesto en la investigación, descubrir los cliques, porque para ello es preciso que la encuesta abarque todos los miembros de la clase.

"Clique", como es sabido, es el subgrupo cerrado de amigos íntimos que se eligen todos ellos unos a otros, recíprocamente.

H) Para estudiar las actitudes políticas de una sociedad cultural que agrupa 5.000 miembros se ha decidido realizar una encuesta por cuestionario a 500 de ellos elegidos arbitrariamente, además de 200 entrevistas a otros tantos socios elegidos al azar, si bien en la realidad los entrevistadores se permitieron sustituir frecuentemente los socios elegidos por otros.

RESPUESTA

Hay en los dos casos, en principio, muestra del universo. Sin embargo, la primera es inadmisible en cuanto no reúne la condición de basarse la elección en un procedimiento racional, si es posible al azar, y además se opone al principio de que dicha elección no debe ser arbitraria, pues hay gran peligro de que prevalezcan criterios subjetivos en ella. En el segundo caso la muestra correcta inicialmente ha resultado viciada en la realidad por la sustitución personal y, por tanto, subjetiva que los entrevistadores se han permitido.

I) Una empresa de mercados para investigar el interés del público mayor de edad de una ciudad donde viven 8.000 blancos, 5.000 negros y 3.000 de otras razas, por una marca de licor determinada, ha decidido realizar una encuesta a una muestra elegida al azar de 400 negros.

RESPUESTA

En este caso también se trata de una muestra inválida, porque aunque el universo es toda la población, sólo se basa en un sector de ella y además se da la circunstancia que es homogéneo, a diferencia de la población, que es heterogénea. Para ser correcta la nuestra debería comprender una parte de las tres razas en la misma proporción. Es decir, debería estar estratificada.

J) Para estimar la proporción de amas de casa que habían visto un programa de TV. en la semana anterior, los investigadores eligieron una muestra al azar de manzanas de la ciudad. En una de dichas manzanas, compuesta de 150 pisos, los encuestadores entrevistaron a 100 amas de casa, que fueron encontradas en el piso al llamar a la puerta.

RESPUESTA

Aquí se tiene que un porcentaje muy elevado de la muestra, 50 de 150, no ha podido ser entrevistado. Ello quita representatividad a la muestra, sobre todo porque muchas de las que faltaban es posible que, por trabajo u otras ocupaciones, estén habitualmente fuera de casa y, por tanto, en ellas la proporción de las que no han visto el programa sea mayor.

38

Supuestos los siguientes estudios que se ha pensado realizar por encuesta muestral, se pide indicar la base y la unidad de la muestra en todos ellos.

A) Análisis de la evolución en sus diversos aspectos de las informaciones religiosas en los periódicos españoles antes y después del Concilio.

RESPUESTA

En primer lugar hay que advertir que el estudio abarca una pluralidad de periódicos, de años y de números de cada periódico dentro de cada año. La muestra general debe comprender, pues, tres submuestras correspondientes a las diversas pluralidades indicadas.

Respecto a los periódicos, la base de la muestra debería ser el registro oficial de todos los periódicos españoles autorizados y la unidad de la muestra cada periódico.

En cuanto a los años, es obvio que la base de la muestra es la lista natural de todos los que abarque la investigación y la unidad cada año.

Asimismo, dentro de cada periódico y año, la base sería el total de números publicados por cada periódico al año y la unidad cada número.

B) Un estudio sobre las condiciones de viviendas familiares de la zona madrileña de Vallecas.

RESPUESTA

En este caso la base de la muestra debería ser el plano de la zona, en el que, numeradas correlativamente por sectores a efectos de muestreo estratificado en su caso, las manzanas constituirían la unidad de la muestra.

C) Un estudio sobre las condiciones estructurales y funcionales de las asociaciones voluntarias privadas de España, con exclusión de las económicas, religiosas, políticas y sindicales.

RESPUESTA

Base de la muestra: El registro oficial de asociaciones. Unidad: cada asociación.

D) Una investigación sobre la relación entre la estabilidad familiar y la clase social en una ciudad pequeña.

RESPUESTA

En este caso habría que recurrir como base de la muestra al censo o padrón de vecinos de la ciudad. La unidad de la muestra sería la familia.

E) Afición al cine de horror y de misterio y tipo de personalidad.

RESPUESTA

En este caso no se ve la posibilidad de obtener como base de la muestra una lista o registro. Por ello habría que adoptar, por ejemplo, el procedimiento de encuestar a uno de cada cinco, ocho, etc., de los asistentes a la proyección de este tipo de películas.

39

En un estudio sociológico sobre una comunidad de 8,500 habitantes mayores de edad se ha elegido una muestra de 850.

Hallar la fracción de muestreo y el coeficiente de elevación. Establecido este último y supuesto que en el estudio realizado han resultado ocho divorciados, determinar el número total de divorciados que es probable existan en la comunidad.

RESPUESTA

Fracción de muestreo: $\dfrac{850 \times 100}{8.500} = 10\ \%.$

Coeficiente de elevación: $\dfrac{8.500}{850} = 10.$

Número probable de divorciados en la población: igual al número de divorciados en la muestra por el coeficiente de elevación: $8 \times 10 = 80$.

40

Se ha proyectado realizar una encuesta sobre el consumo de leche en las familias. El número de familias del universo es 6.000, y la amplitud de la muestra, 840.

Teniendo en cuenta la fracción de muestreo que representa esta muestra, se pide determinar el número de unidades que se deben incluir en la muestra de cada uno de los estratos profesionales siguientes que comprende el universo:

Profesionales liberales. 100.
Comerciantes, 200.
Obreros, 2.000.
Agricultores, 600.
Servicios, 1.900.
Empleados, 1.200.;

RESPUESTA

Fracción de muestreo: $\dfrac{840 \cdot 100}{6.000} = 14\ \%.$

Unidades de muestra para cada estrato:

Profesiones liberales	100.	14 % =	14
Comerciantes	200.	14 % =	28
Obreros	2.000.	14 % =	280
Agricultores	600.	14 % =	84
Servicios	1.900.	14 % =	266
Empleados	1.200.	14 % =	168
TOTAL			840

41

Se ha proyectado realizar una encuesta entre abogados oficialmente en ejercicio de Madrid, con objeto de conocer su opinión sobre la reforma del Código Civil.

Señalar las posibles unidades y bases de la muestra, en el caso de que se estimen puedan ser varias, e indicar razonadamente cuál sería más adecuada a los fines perseguidos.

RESPUESTA

La unidad de la muestra ha de ser, según el enunciado, los abogados en ejercicio, pero hay que advertir que éstos actúan bien individualmente, bien agrupados o bien existen abogados destacados que tienen en su despacho una serie de ayudantes. Por tanto existe la posibilidad de tomar como unidad de la muestra a cada abogado en ejercicio o bien los despachos y bufetes, ya estén formados por uno o varios abogados.

Dado que el objeto de la investigación es conocer la opinión de los abogados en ejercicio y que cada abogado, aunque trabaje en un despacho con otros, es de suponer que tenga su opinión propia, más o menos diferente de la de sus compañeros, parece que la unidad de la muestra a elegir en este caso debería ser cada abogado individual en ejercicio, independientemente de que trabaje con o a las órdenes de otros. Por otra parte, la elección como unidad de la muestra de los despachos plantea problemas más difíciles para formar la base de la muestra.

Determinada de esta forma la unidad de la muestra, para obtener la base de ésta, o lista de todos los abogados en ejercicio de Madrid, se podría recurrir a varios procedimientos, como, por ejemplo, utilizar los registros fiscales o las listas del Colegio de Abogados. Es obvio que es preferible elegir los registros y listas del Colegio de Abogados, en cuanto constituyen el documento oficial de letrados en ejercicio. Nótese que al dirigir la encuesta a los abogados en ejercicio se excluyen por este hecho mismo a los que estén matriculados en el Colegio como abogados sin ejercicio.

42

Se pretende efectuar una encuesta nacional sobre opiniones, actitudes y valores de la juventud española. Según el censo de 1960, el número de jóvenes de quince a veinticuatro años era, en números redondos, de 4.700.000. La fracción de muestreo se ha fijado en 1/1.000.

Según los datos estadísticos que se poseían, la composición por actividades de la juventud española era la siguiente:

Agricultura, 15 por 100.
Industria, 25 por 100.
Comercio, 20 por 100.
Oficinas y servicios, 35 por 100.
Estudiantes, 5 por 100.

El equipo que realiza la encuesta cree que reviste interés especial conocer las opiniones y actitudes de la juventud estudiantil, pero resulta que el número de encuestas a realizar a universitarios, según la proporción que les corresponde, es sólo de 235, número que se considera insuficiente al efecto.

Se pide determinar qué se puede hacer para que, sin variar el tamaño de la muestra, los estudiantes estén representados en ella con un peso cuatro veces superior al que les correspondería según la importancia de su estrato en el universo, indicando la composición resultante de la muestra por estratos.

RESPUESTA

Como se ha indicado en la introducción, es admisible, cuando un estrato no está suficientemente representado en el universo, forzar o aumentar la representación de este estrato en la muestra, disminuyendo en la misma proporción la participación en ella de otros estratos suficientemente representados.

En este caso, como se dice en el enunciado que se estima que los estudiantes deben estar representados con un peso cuatro veces superior al que les corresponde según el universo, 5 por 100, es claro que su proporción en la muestra deberá ser de un 20 por 100. Esto supone un aumento del 15 por 100, que deberá ser reducido en otros estratos, por ejemplo, un 8 por 100 en oficinas y servicios, un 4 por 100 en comercio y un 3 por 100 en la industria.

Según esto, la composición de la muestra por estratos sería:

	%
Agricultura	15
Industria	22
Comercio	16
Servicios y oficinas	27
Estudiantes	20
TOTAL	100

43 (A)

Efectuada la encuesta a que se refiere el ejercicio anterior y recogidos los cuestionarios dentro de cada estrato en la proporción prevista, se han declarado partidarios de un sistema económico socialista los siguientes porcentajes –supuestos– dentro de cada estrato.

	%
Agricultura	40
Industria	70
Comercio	40
Servicios	50
Estudiantes	90
Total nacional	59,30

Teniendo en cuenta la alteración realizada en la composición de la muestra con relación al universo, se pide indicar si los porcentajes de cada estrato reflejan efectivamente la extensión de esta actitud en su estrato y si el porcentaje total nacional la refleja en relación a toda la juventud.

RESPUESTA

Los porcentajes de cada estrato reflejan efectivamente la adhesión a la actitud en cuestión dentro de su estrato, pero no así el total nacional. Este la reflejaría si la composición porcentual de los estratos del universo fuera la misma que la de la muestra. Como según el ejercicio anterior se ha alterado voluntariamente la composición de la muestra, esta alteración implica que en concreto el mayor peso concedido a los estudiantes que precisamente son los que tienen una actitud más favorable al sistema económico socialista suponga un aumento del total nacional respecto al que hubiera resultado si la muestra guardara la misma proporción en los estratos que el universo.

43 (B)

Dada la conclusión anterior, se pide determinar el total nacional efectivo del porcentaje de jóvenes partidarios del sistema económico socialista, corrigiendo la desviación de que se trata mediante la ponderación o multiplicación de los resultados de cada estrato por el porcentaje en el universo y dividiendo el producto por 100.

RESPUESTA

Jóvenes partidarios de un régimen económico socialista:

GRUPOS DE ACTIVIDAD	*% de favorables*	*% del estrato en el universo*	*% ponderado de los favorables en cada estrato*
Agricultura	40	15	6
Industria	70	25	17,5
Comercio	40	20	8
Servicios	50	35	17,5
Estudiantes	90	5	4,5
TOTAL NACIONAL	59,30		53,50

Se ve que el total nacional partidarios del régimen económico socialista, una vez corregida la alteración efectuada en los estratos de la muestra, es el 53,50 por 100, y no el 59,30 por 100, como resultaba en ésta.

Tanto en este caso, de muestras estratificadas alteradas voluntariamente, como en el de que la muestra obtenida después de efectuada la encuesta difiera sensiblemente en algunos estratos de los porcentajes de éstos en el universo, se deben corregir de este modo los resultados totales.

44

En un sondeo sobre las cuestiones socio-económico-políticas que preocupaban a la población belga, se eligió una muestra de manera estratificada, teniendo en cuenta dos cortes o características: a) la región: flamenca, valona y departamento de Bruselas-capital, y b) el tamaño de la comunidad.

Los porcentajes de representación en la muestra de los distintos estratos de cada corte o característica fueron, una vez realizada la encuesta, los siguientes:

	%
Región flamenca	52,20
Región valona	32,4
Bruselas-capital	15,4
	100

	%
De menos de 2.000 habitantes	15
De 2 a 5.000	18,2
De 5 a 10.000	16,6
De 10 a 20.000	15,8
De 20 a 50.000	16,7
De 50 a 100.000	8,6
De 100.000 y más	9,1
	100

Verificar el equilibrio de la muestra respecto del universo, teniendo en cuenta que la representación en éste de cada estrato era, en la época de la encuesta, la que sigue:

	%
Región flamenca	49,6
Región valona	33,7
Bruselas-capital	16,7
	100

	%
Menos de 2.000 habitantes	15,1
De 2 a 5.000	18,5
De 5 a 10.000	17,2
De 10 a 20.000	14,9
De 20 a 50.000	16,7
De 50 a 100.000	8,3
De 100.000 o más	9,3
	100

RESPUESTA

Para verificar el equilibrio pedido es preciso en primer lugar formar un cuadro comparativo de los porcentajes de cada categoría en la muestra y el universo, similar al que sigue:

CATEGORIAS	*% en muestra*	*% en universo*
Región:		
Flamenca	52,2	49,6
Valona	32,4	33,7
Bruselas-capital	15,4	16,7
Tamaño comunidades:		
De menos de 2.000 habitantes	15,0	15,1
De 2 a 5.000	18,2	18,5
De 5 a 10.000	16,6	17,2
De 10 a 20.000	15,8	14,9
De 20 a 50.000	16,7	16,7
De 50 a 100.000	8,6	8,3
De 100.000 y más	9,1	9,3

La simple inspección ocular muestra que no existen variaciones importantes en las proporciones de cada estrato en la muestra y en el universo o población total. No obstante, para comprobar si estas diferencias son significativas o se deben simplemente al azar, se puede utilizar la prueba de χ^2.

(Ejercicio basado en el trabajo de V. Delruelle, J. Coenen y D. Maigray: *Les problèmes que préoccupen les belges.* "Rev. de L'Institut de Sociologie, 1966, 2.)

10. ELECCION DE LA MUESTRA

LA ELECCION DE LA MUESTRA Y SUS REQUISITOS BASICOS

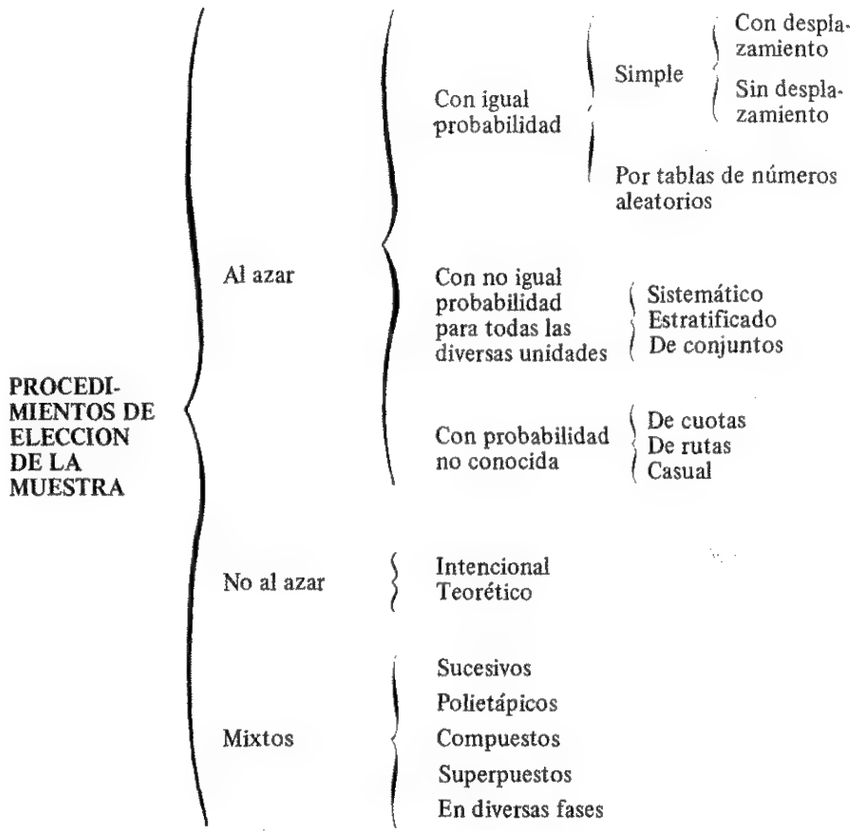
Operativamente la muestra es una elección de unidades dentro de un conjunto, que no es otro que la base de la muestra. Esta es pues el resultado de una elección y por tanto su bondad depende de la bondad de la elección.

La bondad de esta elección depende de dos condiciones fundamentales: una estadística y otra teórica. De acuerdo con la primera, debe ser válida la generalización de los resultados obtenidos en la muestra a la población. Este requisito se logrará si la amplitud de la muestra es estadísticamente proporcionada a la magnitud de la población y en su composición por estratos es representativa o reflejo fiel de la población.

Según la segunda, la muestra elegida debe ser adecuada para el logro de los objetivos de la investigación y la prueba de las hipótesis teóricas que constituyan su razón de ser. Como señala Galtung (1973, 52) para poner a prueba en una investigación la hipótesis sustantiva "a mayor edad, mayor impacto de la renta en la actitud hacia la gran empresa", es obvio que se necesita una muestra que incluya gente joven y de edad, con altos y bajos ingresos, a favor y en contra de la libre empresa. Esta condición implica que la muestra sea heterogénea, es decir, contenga individuos con niveles altos y bajos en las variables que forman las hipótesis. Si una muestra es estadísticamente representativa de la población contendrá normalmente casos de individuos con valores extremos y medios en las variables investigadas, salvo en el caso de hipótesis o análisis complicados.

Estadísticamente, el principio básico de elección de la muestra es que ésta se haga, siempre que sea posible, de tal modo que cada elemento de la población tenga la misma probabilidad de ser elegido.

Este requisito se cumple si la elección tiene lugar por un procedimiento aleatorio riguroso. Pero no siempre es posible realizarlo así, de ahí que existan diversos procedimientos de elección de la muestra, que se pueden clasificar, como se hace en el cuadro que sigue, según se conozca o no la probabilidad de elección de cada unidad y, en caso afirmativo, según sea esta probabilidad igual o distinta para todas las unidades de la población.



El conocimiento de la probabilidad de elección implica la existencia de una seguridad conocida estadística, por tanto no absoluta sino razonablemente probable, de que la muestra es representativa de la población, mientras en las no probabilísticas dicha seguridad estadística no

existe y no se conoce, por tanto, la garantía de ser representativa del universo ofrecida por la muestra obtenida.

AZAR SIMPLE Y AZAR SISTEMATICO

Azar simple.— Consiste en sacar, por medio de un sorteo riguroso, una serie de unidades de un universo, hasta completar el tamaño de la muestra que hayamos fijado.

Obviamente este sistema no presenta ninguna dificultad, salvo que se puede o no insacular, o meter en el bombo de nuevo, las unidades que se vayan sacando, lo cual influye en la probabilidad de salir las demás en las restantes extracciones. Si se meten, la elección se llama con reemplazamiento y sin reemplazamiento en el caso contrario.

Según Nam Lin (1976, 147), la muestra con reemplazamiento es el tipo de muestreo ideal según la teoría de las muestras. Sin embargo, el muestreo sin reemplazamiento no presenta tantas dificultades prácticas y es de hecho el más usado.

Este procedimiento aunque es simple, también presenta sus complicaciones, sobre todo en una población o universo muy amplio. Es preciso tener bombo, bolas suficientes, comprobarlas antes de meterlas en el bombo, que estén todas, sacarlas una por una.

Por eso es bastante corriente usar en su lugar tablas de números de azar. Estas tablas son simplemente listas de números sacados al azar o por sorteo. Son conocidas y están editadas en España las de Fisher y Yates (Ed. Aguilar).

Estas tablas se pueden utilizar de diversas maneras. Una de ellas consiste en ir eligiendo de seguido, empezando por una columna de la tabla determinada, números inferiores al universo hasta completar los que necesitamos para la muestra.

Si se da el caso, tanto en este sistema como en el anterior, de que un mismo número salga repetido este no se toma en cuenta y se saca o toma otro en su lugar.

En síntesis, de acuerdo con B. Lazerswitz (1968, p. 279), la elección de una muestra por el sistema de azar simple comprende los siguientes pasos:

1) Listar todos los elementos de la población y asignarles números consecutivos del 1 al N base.

2) Decidir el tamaño de la muestra deseado, previa fijación del nivel de confianza y el error.

3) Usar una tabla de números aleatorios y elegir diferentes números que caigan entre el 1 y N.

4) Los elementos listados en la población que tienen asignados estos varios números al azar, forman la muestra.

Azar sistemático.— Con este sistema, se simplifica la elección aún más que con las tablas. En él sólamente se elige al azar o por sorteo una primera unidad de la muestra inferior al coeficiente de elevación.

Las restantes se obtienen sumando a dicho número el coeficiente de elevación sucesivamente, hasta obtener todos los números necesarios para la muestra.

Este procedimiento, para que se pueda aplicar exige que la base de la muestra no presente ninguna ordenación, por ejemplo de 10 en 10, de 15 en 15, etc., que coincida con el coeficiente de elevación.

Otro problema en su aplicación, es el caso de que el coeficiente de elevación no sea un número entero. Se podrían despreciar los decimales, pero habría que tener en cuenta que entonces una parte última del universo, más o menos extensa, no tendría ninguna posibilidad de salir en la muestra. Esto se puede solucionar utilizando unas veces como coeficiente de elevación el indicado, despreciando los decimales, y otras, el número inmediato superior de modo que se abarque todo el universo.

Un problema que presenta este muestreo es el modo de sustituir las unidades llamadas vacías: individuos que no se encuentran, viviendas no ocupadas, etc. En el azar simple, se pueden sustituir efectuando las nuevas elecciones que sean necesarias. En el azar sistemático se descarta este procedimiento, aunque se podría utilizar subsidiariamente.

Una solución a este problema puede ser realizar una nueva elección sistemática para todas las unidades vacías, en la que el coeficiente de elevación sería el que resultara de dividir el universo por el conjunto de las unidades vacías.

Además de este procedimiento de muestreo sistemático, que es el más recomendable por ser más riguroso, existen también otras maneras de obtener la muestra, que se consideran también como muestreos sistemáticos.

Entre ellos se pueden mencionar los siguientes:

1) Formación de la muestra con todos los individuos cuyo nombre empiece por una palabra del alfabeto determinada, por ejemplo la D o la S.

2) Formación de la muestra con los que hayan nacido en determinados días del año, como por ejemplo el 15 de enero y el 8 de julio.

3) Elección por la última o las dos últimas cifras del número que tengan asignados los elementos de la muestra en la base de ésta. Por ejemplo, se eligen todos los números que terminen en 5 o en 72. (Ver ejercicios 45, 46).

MUESTREO ESTRATIFICADO Y POR CUOTAS

Cuando el universo no es naturalmente homogéneo, sino que está formado por estratos diferentes, que constituyen categorías importantes para la investigación, la elección de la muestra no debe hacerse globalmente, para todos los estratos de una vez, porque entonces nos expondríamos a que unos estratos estuvieran más o menos representados que lo proporcionalmente debido.

En este caso caben tres tipos de muestreo que se llaman de Bernoulli, Lexis y Poisson.

En el primero, de Bernoulli, se hace el muestreo sin tener en cuenta los estratos. En el segundo, sólo se muestrea un estrato en representación de todos.

En el 3°, que es el más recomendable si se posee una base de la muestra por estratos, se realiza una elección de la muestra independiente para cada estrato, por el procedimiento que nos parezca oportuno de los indicados anteriormente: sorteo, tablas, sistemático.

Según Kish, la muestra estratificada se utiliza por tres tipos de razones:

a) Reducir la variación de los resultados de la muestra respecto a la población total y obtener una mayor precisión en las estimaciones;

b) Usar diferentes métodos de muestreo y fracciones de muestreo diversas en los distintos estratos de población;

c) Acomodarse a subpoblaciones que representan dominios determinados de estudio.

Por otra parte, los rasgos base de la estratificación deben relacionarse con las variables que se estimarán en la investigación. Se debe procurar

que las unidades de muestreo dentro de cada estrato sean lo más homogéneas posible.

Si no existe ni se puede formar una base de la muestra por estratos, pero sabemos la composición por estratos del universo en tantos por cientos y la encuesta se realiza por entrevista, se recurre al procedimiento que se llama muestreo por cuotas. Consiste en asignar a cada entrevistador un número de entrevistas a realizar, indicándole las que tienen que corresponder a cada estrato, y dejando a su arbitrio la elección concreta de las unidades de la población a entrevistar.

Como se trata de un sistema imperfecto, el error que supone la muestra será mayor, por lo que se ha estimado empíricamente que el tamaño de la muestra practicada por cuotas, deberá ser mayor en un 50 % al de la muestra elegida al azar, para que los errores sean equivalentes.

Según señala Mayntz (1969, p. 108) el muestreo por cuotas encuentra su fundamento teórico en el hecho de que si se conocieran todas las características del universo y sus proporciones correspondientes, sería posible organizar una muestra que estuviera "cuotificada" en todas sus dimensiones, de tal manera que fuera representativa del universo, sin necesidad de una extracción aleatoria individual de sus unidades.

Este procedimiento de muestreo presenta los inconvenientes siguientes:

1) No pueden aplicarse las fórmulas estadísticas para hallar, por ejemplo su error.

2) Existe el peligro de que los entrevistadores elijan por comodidad a personas cercanas a ellos, de su "entorno social". También puede ocurrir que para encontrar a personas a quien entrevistar pregunten a aquéllos a quienes han sido ya entrevistados. Es lógico suponer que estos probablemente indicarán también a personas de su mismo entorno social, y éstas a otras, con lo que se produce el efecto llamado de la "bola de nieve".

Para obviar estos inconvenientes, se ha ideado el muestreo de rutas, que consiste en fijar a cada entrevistador un itinerario definido en todos sus detalles, indicándole exactamente en qué puntos debe realizar cada entrevista.

Cuando los estratos pertenecen a dos o más tipos de categorías, se dice que la muestra tiene dos o más cortes, como por ejemplo, en el caso de que se considere una población estratificada, según la profesión y la edad. (Ver ejercicios 47, 48)

MUESTREO DE CONJUNTOS Y DE EXTRACCION SUCESIVA O POLIETAPICO

Este tipo de muestreo se refiere al caso en el que las unidades de la muestra no sean simples, sino compuestas o colectivas, es decir, de talla superior a la unidad.

El muestreo de los conjuntos *o clusters* en inglés, se puede realizar por algunos de los procedimientos básicos indicados, sorteo, tablas, azar sistemático.

Una vez obtenidos los conjuntos muestreados a observar, se puede realizar la encuesta a todos los individuos que comprenda cada cluster. Igualmente se puede efectuar otra nueva muestra con referencia a cada cluster para elegir unidades que representen al cluster. A esto es lo que se llama muestreo de extracción sucesiva o por etapas, que pueden ser más de dos.

Según H. Blalock la ventaja de este tipo de muestreo consiste en que ahorra la dificultad y el gasto de obtener censos completos y concentra especialmente los puntos de entrevista. (Ver ejercicios 50, 51).

OTROS TIPOS DE MUESTREO

Se llaman muestras superpuestas, a dos muestras distintas obtenidas de la misma base de la muestra mediante la realización de dos sorteos o dos sistemas de muestreo distintos.

Este procedimiento es muy útil cuando se quiere lograr una valiosa corroboración o verificación de los resultados obtenidos en una encuesta.

Muestreo teorético.— Es el orientado directamente a la prueba de hipótesis. En él se parte de la previsión anticipada de las tablas a formar con los datos y los análisis a efectuar para dicha prueba. El muestreo entonces se dirige a obtener el número necesario de todos los tipos de datos exigidos para cubrir las diferentes casillas de la tabla. Esta clase de muestreo puede no permitir la generalización a una población más amplia que la muestra.

Muestras sucesivas.— Los fenómenos sociales evolucionan con el paso del tiempo. Sin embargo, es claro que una encuesta sólo nos proporciona información sobre el momento en que se realiza, pero no sobre la evolución posterior del fenómeno que se estudia.

Por eso, en el caso de que nos interese conocer la evolución del fenómeno estudiado con el paso del tiempo, por ejemplo, estudios de mercados y estudios electorales, la solución es repetir la encuesta a nuevas muestras cada cierto período de tiempo. Esto es lo que se llama muestreo sucesivo o muestras sucesivas y estudios de panel, cuando el grupo encuestado es el mismo en las sucesivas encuestas.

En el caso de tener que realizar un muestreo de este tipo, se puede optar por algunos de los siguientes procedimientos básicos.

1) Hacer cada intervalo de tiempo una muestra nueva completa. Acaso sería el procedimiento más exacto, pero es también el sistema más costoso y difícil.

2) Utilizar la misma muestra inicial en todos los sucesivos sondeos o encuestas. Este procedimiento es económico, pero presenta la dificultad de la posible desaparición de elementos de la muestra.

3) En lugar de variar toda la muestra, o conservarla toda, se puede seguir el camino intermedio de renovarla sólo en su mitad, tercera y cuarta parte, etc.

4) Otro procedimiento es el que se llama de "rotación de la muestra" que en principio parece el más aconsejable.

En este procedimiento se forman dos muestras distintas y cada una de ellas se divide en un número determinado de submuestras, ocho por ejemplo.

Las submuestras de cada muestra se van combinando en cada aplicación de la muestra siguiendo un orden rotatorio. En la encuesta sobre población americana donde se utiliza este procedimiento, cada subgrupo de la muestra se incluye durante 4 meses.

Este procedimiento de rotación de la muestra tiene las siguientes ventajas:

1. Evita el cansancio de los investigadores de entrevistar siempre a las mismas personas.

2. Impide que los encuestados adquieran resabios y rutinas al tener que contestar siempre a las mismas preguntas.

3. Aleja el peligro de que los entrevistadores puedan ser influidos por la repetición en la forma de realizar las entrevistas.

Por último, también es admisible realizar una muestra inicial completa y en los sondeos sucesivos trabajar con una submuestra, o muestra extraída de la muestra inicial, y por tanto, bastante más reducida que ésta. Es claro que ésta submuestra es más fácil de construir que otra muestra distinta, y menos costosa de llevar a efecto.

Muestras compuestas.— Se llaman así, los sistemas de muestreo que comprenden dos o más muestras distintas, según las diferentes partes del fenómeno estudiado.

Por ejemplo, en un estudio que abarque una parte rural y otra urbana, se podría aplicar probablemente un distinto tipo de muestreo para las zonas urbanas y las rurales. Por ejemplo, al azar con base en los censos o listas para las primeras y el de clusters o conjuntos para las zonas rurales.

Muestreo opinativo o estratégico.— En el caso sobre todo de muestras de conjuntos, y en general cuando los elementos de la muestra son muy reducidos y diversos, puede ser aconsejable utilizar este tipo de muestreo, opinativo, o estratégico, en el que en vez de realizar la elección de las unidades de la muestra al azar, se realiza razonadamente por los investigadores con los asesoramientos que sean precisos. (Ver ejercicio 52)

Muestreos de diversas fases.— Según M. Grawitz (1973, Vol. 1, pág. 73), en este caso hay combinación de varios sondeos, con variación de las fracciones de sondeo, aunque las unidades siguen siendo las mismas.

Supongamos que debe efectuarse una encuesta relativa al consumo de una muestra de matrimonios. En este caso un sondeo "normal" contendrá sólo un reducido número de matrimonios con recursos muy altos o muy bajos. Por ello sería aconsejable realizar el sondeo en dos fases, conservando el matrimonio como unidad de sondeo.

En la primera fase se extrae una gran muestra de matrimonios, a los que se les somete a una rápida encuesta, en la que cada pareja indica el total de sus ingresos. En la segunda fase esta gran muestra es estratificada según el importe indicado de sus ingresos.

En los estratos extremos se somete íntegramente la población al cuestionario. En los estratos intermedios se aplica según se precise a fracciones más o menos reducidas.

Muestras compensadas.— Esta técnica tiene aplicación cuando se conoce el valor medio que alcanza una de las características estudiadas en el universo. Como si por ejemplo, tuvieramos que obtener una muestra de una población y conociéramos de antemano la media del peso de los individuos que la componen.

Entonces, obtenida la muestra, es muy poco probable que la media de sus pesos fuera igual a la media conocida del universo.

Sin embargo, lo que se puede hacer, y en esto consiste la compensación de la muestra, es sustituir unidades de la muestra por otras nuevas elegidas también al azar, en el caso de que sus valores acerquen más la media de la muestra a la del universo, pudiendo llegarse incluso por sucesivas sustituciones a hacer que coincidan ambas medias. (Ver ejercicio 49)

EJERCICIOS

45

De una población de 500 unidades se quiere obtener una muestra de 25 para realizar una encuesta piloto.

Hallar la muestra deseada por medio del procedimiento aleatorio simple, utilizando la tabla de números al azar inserta en la página siguiente, y sin tablas.

RESPUESTA

Utilizando la hoja adjunta de las tablas de Fisher y Yates, y empezando por la segunda columna, los números al azar de la muestra, por orden de aparición, serían:

369, 428, 385, 495, 169, 094, 441, 403, 332, 485, 060, 333, 423, 339, 150, 223, 099, 469, 425, 329, 050, 316, 173, 491, 074.

Sin tabla es preciso disponer de una bolsa o bombo con 500 bolas numeradas del 1 al 500 o realizar un sorteo por cada una de las cifras del universo entre los diez primeros números, del 0 al 9.

El procedimiento, como es sabido, es ir extrayendo bolas al azar hasta completar las 25. Las bolas extraídas, una vez anotadas, es preferible insacularlas de nuevo, para que no se altere la probabilidad de las restantes extracciones.

46

En el supuesto a que se refiere el ejercicio anterior, hallar la muestra deseada utilizando el procedimiento aleatorio sistemático.

Criticar la validez de este procedimiento.

RESPUESTA

El coeficiente de elevación es: $\frac{500}{25} = 20$.

Adoptando el primer número de la primera columna de las tablas aleatorias inferior a 20, 003, tendremos la siguiente muestra:

3 + 20 = 23 + 20 = 43 + 20 = 63, 83, 103, 123, 143, 163, 183, 203, 223, 243, 283, 303, 323, 343, 363, 383, 403, 423, 443, 463 y 483.

La validez de este procedimiento depende de que el universo no esté ordenado en la base de la muestra de 20 en 20 o en alguna forma que pueda dar lugar a que todas o algunas de las unidades elegidas sean similares en algún sentido y no representativas de todas las características de la población estudiada.

47

Se ha proyectado realizar una encuesta entre los camioneros españoles sobre la marca de neumáticos que utilizan, sobre su satisfacción con ella, y los problemas que tienen con sus neumáticos.

Se pide indicar el procedimiento más adecuado para elegir la muestra y las operaciones fundamentales que habría que efectuar para realizar la elección.

Téngase en cuenta que los camioneros, por razón de su oficio, sólo muy ocasionalmente se encuentran en sus domicilios.

RESPUESTA

Aunque se tuviera o pudiera formar una base de la muestra, no se podría emplear ningún procedimiento aleatorio por sorteo, de elección de la muestra, ya que serían muy pocos los camioneros que se podrían encontrar en sus domicilios.

Habría que utilizar, por tanto, un procedimiento al azar imperfecto, que podría ser similar al de hap hazard, determinando un conjunto de puntos neurálgicos en las carreteras españolas (restaurantes) y entrevistar a los que se encontraron en dichos puntos.

48

Se ha proyectado realizar una investigación sociológica sobre movilidad social intergeneracional —de los hijos respecto de sus padres—, a una muestra estratificada según dos cortes o características, una la ocupación (profesiones liberales, funcionarios y empleados) y otra la edad (de treinta a cincuenta años y de cincuenta a setenta) en una capital de provincia. El tamaño de la muestra determinado es de 500: de ellos 50 de profesiones liberales, 150 funcionarios y 300 empleados. En cuanto a la edad, 280 deberán tener de treinta a cuarenta y nueve años, y 220, de cincuenta a setenta. Estos valores corresponden a las dimensiones respectivas del universo.

Teniendo en cuenta que no existe base de la muestra, indicar el procedimiento que corresponde aplicar para la selección de los individuos de cada estrato a encuestar y confeccionar un cuadro al efecto para que realice cincuenta entrevistas un entrevistador.

TABLA DE NUMEROS ALEATORIOS

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
1	03	47	43	73	86	36	96	47	36	61	46	98	63	71	62	33	26	16	80	45	53	74	23	99	67	61	32	28	69	84	94	62	67	86	24	98	33	41	19	95
2	97	74	24	67	62	42	81	14	57	20	42	53	32	37	32	27	07	36	07	51	63	38	06	86	54	99	00	65	26	94	02	82	90	23	07	79	62	67	80	60
3	16	76	62	27	66	56	50	26	71	07	32	90	79	78	53	13	55	38	58	59	35	30	38	21	46	06	72	17	10	94	25	21	31	74	96	49	28	24	00	49
4	12	56	85	99	26	96	96	68	27	31	05	03	72	93	15	57	12	10	14	21	65	43	36	92	69	65	51	18	37	88	61	38	44	12	45	32	92	84	83	65
5	55	59	56	35	64	38	54	82	46	21	31	62	43	90	90	08	18	44	32	53	98	25	37	55	26	01	91	82	81	46	74	71	12	94	97	24	02	71	37	07
6	16	22	77	94	39	49	54	43	54	82	17	37	93	23	78	87	35	20	96	43	03	92	18	27	46	57	99	16	96	56	30	33	72	85	22	84	64	38	56	98
7	84	42	17	53	31	57	24	55	06	88	77	04	74	47	67	21	76	33	50	25	64	55	22	21	82	48	22	28	06	00	61	64	13	54	91	82	78	12	23	29
8	63	01	63	78	59	16	95	55	67	19	98	10	50	71	75	12	86	73	58	07	85	07	26	13	89	01	10	07	82	04	59	63	69	38	03	69	11	15	83	80
9	33	21	12	34	29	78	64	56	07	82	52	42	07	44	38	15	51	00	13	42	58	54	16	24	15	51	54	44	82	00	62	61	65	04	69	38	18	65	18	87
10	57	60	86	32	44	09	47	27	96	54	49	17	46	09	62	90	52	84	77	27	34	85	27	84	87	61	48	64	56	26	90	18	48	13	26	37	70	15	42	57
11	18	18	07	92	46	44	17	16	58	09	79	83	86	19	62	06	76	50	03	10	72	63	31	17	69	71	50	80	39	56	38	15	40	11	48	43	40	45	86	98
12	26	62	38	97	75	84	16	07	44	99	83	11	46	32	24	20	14	85	88	45	02	30	27	59	37	62	75	41	66	48	86	97	80	61	45	23	53	04	01	07
13	23	42	40	64	74	82	97	77	77	81	07	45	32	14	08	32	98	94	07	72	05	45	93	15	22	60	21	75	46	91	98	77	27	85	42	28	88	61	08	84
14	52	36	28	19	95	50	92	26	11	97	00	56	76	31	38	80	22	02	53	53	07	08	55	18	04	45	44	73	13	90	24	94	96	61	02	57	55	66	83	15
15	37	85	94	35	12	83	39	50	08	30	42	34	07	96	88	54	42	06	87	98	01	85	89	05	66	51	10	19	34	88	15	84	97	19	75	12	76	39	43	78
16	70	29	17	12	13	40	33	20	38	26	13	89	51	03	74	17	76	37	13	04	72	84	71	14	35	19	11	58	49	26	50	11	17	17	70	86	31	57	20	18
17	56	62	18	37	35	96	83	50	87	75	97	12	25	93	47	70	33	24	03	54	98	78	28	16	84	13	52	53	94	53	75	45	69	30	96	73	89	65	70	31
18	99	49	57	22	77	88	42	95	45	72	16	64	36	16	00	04	43	18	66	79	45	17	75	65	57	28	40	19	72	12	25	12	74	75	67	60	40	60	81	19
19	16	08	15	04	72	33	27	14	34	09	45	59	34	88	49	12	72	07	32	45	96	76	28	12	54	22	01	11	94	25	71	96	16	16	88	68	64	58	74	45
20	31	16	93	32	43	50	27	89	87	19	20	15	37	00	49	52	85	66	60	44	43	31	67	72	30	24	02	94	08	63	38	32	36	66	02	69	36	38	25	39
21	68	34	30	13	70	55	74	30	77	40	44	22	78	84	26	04	33	46	09	52	50	44	66	44	21	66	05	58	04	62	68	15	53	35	17	42	36	48	96	32
22	74	57	25	65	76	59	29	97	08	60	71	91	38	67	54	13	58	18	25	27	22	66	22	15	86	26	63	74	41	99	58	42	36	84	82	58	37	52	13	51
23	27	32	37	86	53	48	55	90	65	72	96	57	69	36	10	96	40	92	42	45	96	24	40	14	51	23	22	30	88	57	95	67	47	21	96	94	69	42	06	07
24	90	39	68	29	61	66	37	32	30	74	77	84	57	03	29	10	45	65	04	26	11	73	91	61	19	60	20	72	93	48	98	57	07	23	69	65	95	39	69	58
25	20	94	98	94	24	68	49	69	10	82	53	75	91	93	30	34	25	20	57	27	78	60	73	99	84	43	89	94	36	34	56	69	47	07	41	90	22	91	07	12
26	16	90	82	66	59	83	62	64	11	12	67	19	00	71	74	60	47	21	29	68	84	37	90	61	56	70	10	23	98	05	85	11	34	76	60	76	48	45	34	60
27	11	27	94	75	06	06	09	19	74	66	02	94	37	34	02	76	70	90	30	86	36	67	10	08	23	98	93	35	08	86	99	29	76	29	81	33	34	91	58	93
28	35	24	10	16	20	33	32	51	16	38	79	78	45	04	91	16	92	53	56	16	07	28	59	07	48	89	64	58	89	75	83	85	82	27	89	30	14	78	56	27
29	38	23	16	86	38	42	38	97	01	50	87	75	60	81	41	40	01	74	91	62	55	19	68	97	65	03	73	52	16	56	00	53	55	90	27	33	42	39	38	87
30	31	96	25	91	47	96	44	33	49	13	34	96	82	53	91	00	52	43	48	85	10	15	83	87	60	79	24	31	66	56	21	48	24	06	93	91	98	94	05	49
31	66	67	40	67	14	64	05	71	95	86	11	05	65	09	68	76	83	20	37	90	53	81	29	13	39	35	01	20	71	34	62	33	74	82	14	53	73	19	09	03
32	14	90	84	45	11	75	73	88	05	90	52	27	41	14	86	22	98	12	22	08	51	86	32	68	92	33	98	74	66	99	40	14	71	94	58	35	94	19	38	81
33	68	05	51	18	00	33	96	02	74	19	07	60	62	93	55	59	33	82	43	90	35	91	70	29	13	80	03	54	07	27	96	94	78	32	66	50	95	52	74	33
34	20	46	78	73	90	97	51	40	14	02	04	02	33	31	08	39	54	16	49	36	37	71	67	95	13	20	02	77	95	94	64	85	04	05	72	01	32	90	76	14
35	64	19	58	97	79	15	06	15	93	20	01	90	10	75	06	40	78	78	89	62	93	60	13	27	83	27	92	79	64	64	72	28	54	96	53	84	48	14	52	98	94
36	05	26	93	70	60	22	35	85	15	13	92	03	51	59	77	59	56	78	06	83	02	96	06	45	64	13	05	00	41	84	93	07	54	72	59	21	45	57	09	77
37	07	97	10	88	23	09	98	42	99	64	61	71	62	99	06	51	29	16	93	15	49	83	43	48	36	92	88	33	69	96	72	36	04	19	76	47	45	15	18	60
38	68	71	86	85	85	54	87	66	47	54	73	32	08	11	12	44	95	92	63	16	84	60	71	62	46	40	80	81	30	37	34	39	23	04	38	25	15	35	71	30
39	14	65	52	68	74	87	37	78	22	41	26	78	63	06	55	13	08	27	01	50	15	17	30	88	71	44	01	14	88	47	89	23	30	63	15	56	34	20	47	89
40	17	53	77	58	71	71	59	36	50	72	12	41	94	96	26	44	95	27	36	99	79	69	10	61	78	71	32	76	95	62	87	00	22	58	40	92	54	01	75	25

Procedente de FISHER et YATES, *Statistical Tables*. (Hay edición en castellano publicada por la Edit. Aguilar.)

RESPUESTA

a) El procedimiento de selección que corresponde aplicar en este caso es el de cuotas, en el que se asigna a cada entrevistador un número determinado de encuestas a realizar con indicación solamente de las que dentro de la cifra general deben corresponder a cada estrato de la muestra.

b) El cuadro a entregar al entrevistador para realizar 50 entrevistas podría ser similar al siguiente:

Entrevistador: D. ...

Número de entrevistas a realizar: 50.

	De 30 a 50 años	De 50 a 70 años	TOTAL
Profesionales	3	2	5
Funcionarios	9	6	15
Empleados	16	14	30
TOTALES	28	22	50

49

Se está realizando una investigación que guarda relación con el coeficiente de inteligencia (Q.I.) de una población cuyo valor medio se conoce y es igual a 11,6 (cifra supuesta). Una muestra previamente sacada al azar de la población ha dado los siguientes Q.I.:

9,1, 13,4, 11,8, 10,3 y 12

Compensar la muestra obtenida, teniendo en cuenta: a) que los valores del Q.I. de las sucesivas unidades extraídas al azar de la población a efectos de la compensación pretendida son las siguientes:

14,2, 10,5, 11,3, 11,8, 13,1, 6,5, 10,7, 9,9 y 11,2

b) Que las operaciones a efectuar son las siguientes:

1. Se halla la media de las unidades de la muestra previamente obtenida. Se ve si coincide con la media del universo. Si coincide, no es necesario compensar; en otro caso, sí.

2. Se sustituye la primera unidad de la muestra previa por la primera unidad obtenida a efectos de la compensación

3. Se halla la media de la muestra así rectificada.

4. Si la media obtenida después de la sustitución se aproxima más a la media del universo, se confirma la sustitución; en otro caso, se rechaza.

5. Realizada la primera sustitución, se siguen sustituyendo de la misma forma las unidades de la muestra primera con las obtenidas posteriormente y así se puede continuar indefinidamente si se quiere hasta que la media de la muestra coincida con la media del universo.

RESPUESTA

$$\text{Media de la muestra previa } \frac{9{,}1 + 13{,}4 + 11{,}8 + 10{,}3 + 12}{5} = 11{,}3.$$

$$\text{Media de la primera sustitución } \frac{14{,}2 + 13{,}4 + 11{,}8 + 10{,}3 + 12}{5} = 12{,}3.$$

La nueva media se aparta más de la media del universo 11,6 que la inicial; luego la sustitución realizada se rechaza.

$$\text{Media de la segunda sustitución } \frac{10{,}5 + 13{,}4 + 11{,}8 + 10{,}3 + 12}{5} = 11{,}6.$$

La nueva media obtenida es igual a la del universo, luego la muestra está ya perfectamente compensada y no es necesario efectuar nuevas sustituciones.

50

En una provincia española que tiene 540 municipios de menos de 1.000 habitantes, se quiere obtener una muestra al azar de 25 para estudiar los hábitos alimenticios rurales

Se pide: 1. Determinar qué tipo de muestreo se trata. 2. Indicar el procedimiento para llevarlo a efecto. 3. Elegir la muestra por medio de tablas al azar empezando por las tres primeras cifras de la primera columna.

RESPUESTA

1. Se trata del tipo de muestreo que los americanos llaman de "clusters", en racimos, y que se puede traducir por conjuntos o conglomerados en el que la unidad de la muestra es de carácter colectivo, ya que en un municipio viven más de un habitante.

2. El procedimiento para llevarlo a efecto es, en principio, el mismo que si la unidad de la muestra fuera individual. Por ello se ha de formar la base de la muestra con todos los pueblos numerados individualmente.

3. Utilizando la tabla y columna indicadas, tendríamos los siguientes números de municipios:

034, 167, 125, 162, 332, 181, 266, 234, 523, 378, 160, 311, 273, 003, 299, 169, 112, 352, 382, 319, 149, 204, 052, 079, 146.

51

En el ejercicio anterior, una vez obtenida la muestra y dado el número de habitantes de los municipios obtenidos, ante la imposibilidad de encuestar a todos ellos, los organizadores han decidido elegir en cada pueblo una de cada diez familias, y dentro de cada familia, uno de sus miembros mayores de dieciséis años.

Se pide: 1. Determinar qué tipo de muestreo se trata. 2. Describir el método de elección de la muestra a emplear.

RESPUESTA

1. Este tipo de muestreo es el llamado de extracción sucesiva o por etapas, pues primero se obtuvo la muestra de Ayuntamientos; luego dentro de éstos hay que obtener una muestra de familias, y dentro de las familias elegir uno de sus miembros.

2. El procedimiento a emplear, de acuerdo con el enunciado, para elegir la muestra, previa la formación de su base numerando las familias del pueblo, sería el de azar sistemático, en el que se elegiría por sorteo una familia de las diez primeras de la lista y después se iría sumando a esta sucesivamente diez. La elección de los miembros de la familia se debería hacer por simple sorteo o según algún criterio lógico, pues, dado el pequeño número de ellos, no tienen aplicación práctica las tablas ni el muestreo sistemático.

52

Se quiere realizar un estudio socioeconómico sobre la banca española a base de una muestra representativa previamente seleccionada.

Se pide: 1. Indicar el procedimiento de muestra a emplear. 2. Señalar el modo de efectuarlo.

RESPUESTA

Dado el pequeño número relativamente de bancos dentro de cada grupo y la gran diversidad de éstos, no sería aconsejable realizar la selección de la muestra por los procedimientos al azar, pues se expondría uno, por ejemplo, a que no saliera ningún gran banco a pesar de su importancia.

Por esto parece sería aconsejable en este caso y otros similares el método de elección razonada, llamado también opinático o estratégico, en el cual los organizadores de la encuesta, con los asesoramientos necesarios, elegirán, dentro de cada tipo de bancos, los que habrían de formar parte de la muestra, teniendo en cuenta todos los fines de la investigación.

11. EL ERROR MUESTRAL

NOCION

En una muestra podemos seguir dos caminos, o bien, calcular el tamaño de la muestra aplicando la fórmula, previa fijación del error que admitimos, o bien, establecer sin más el tamaño y luego calcular el error aplicando la fórmula de éste. Matemáticamente, la fórmula del tamaño se deriva de la del error. Por eso estudiaremos ésta primero.

Todo error lo es en relación a algún patrón o punto de referencia. En el caso del error muestral, el punto de referencia es la población de la que se obtiene o saca la muestra.

La muestra, como es sabido, no es el universo, sino una representación en pequeño de él. Por perfecta que resulte, habrá normalmente una divergencia entre los valores medios estadísticos obtenidos de la muestra con relación a alguna variable y los derivados del universo o población. Por ejemplo, si según una muestra resulta que una media de un 25 % de las personas consultadas asiste con frecuencia semanal al cine, si se hubiera hecho la encuesta a toda la población se habría obtenido seguramente una media distinta en mayor o menor grado. Precisamente en esta divergencia entre los valores medios obtenidos de las muestras y los valores medios del universo, consiste el error muestral, que podemos llamar real.

La única manera de poder conocer este error muestral real es realizar la encuesta a la vez que a la muestra a toda la población. Por ello, hay que descartar, en la práctica, el conocimiento de este error muestral real, porque al exigir realizar la encuesta a toda la población y no solo a la muestra, por este hecho mismo la muestra pierde su razón de ser.

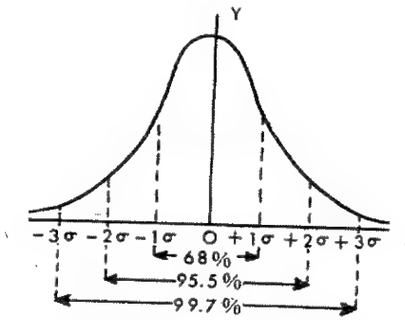
Cuando hablamos, pues, del error muestral no nos referimos a este error muestral real, no conocido, sino a un error muestral determinado estadísticamente, de tipo genérico, válido para todas las muestras posibles, del mismo tamaño.

De ahí que este error muestral, estadístico, sirva para darnos no un error específico y determinado, sino únicamente para establecer el intervalo de confianza dentro del que nos movenos en la muestra, es decir, los límites formados por la media de la muestra, más o menos el error en cuestión, dentro del que se debe encontrar la media del universo, con el grado de probabilidad con que se trabaje, dos o tres sigmas.

Por ejemplo, si en una investigación ha resultado que la muestra nos dice que un 20 % de la población tiene estudios superiores y el error muestral es del 2 % , este error indica que la media del universo de los que tienen estudios superiores debe hallarse con una probabilidad del 95,5 % si hemos trabajado al nivel de confianza de dos sigmas, entre el intervalo de seguridad de 20 ± 2, o sea, entre el 18 y el 22 %.

Según todo esto, está claro que el error muestral, no indica si es por ejemplo del 2 % , que todo resultado de cualquier muestra presente una desviación del 2 % , en relación a los valores respectivos del universo.

Lo que señala, es que estos valores del universo deben encontrarse en el intervalo de confianza o espacio cuantitativo formado por la media, el porcentaje obtenido, más o menos el 2 %.



Dentro, pues, del intervalo, en cualquier punto o valor del mismo es donde deben encontrarse los valores del universo, pero no con una seguridad total, sino con el grado de probabilidad del dos o tres sigmas que se haya elegido para la muestra.

El error muestral, pues, no representa otra cosa que el error estadístico máximo de la muestra, y además genérico, en cuanto es, como he indicado, válido para el conjunto de todas las distintas muestras del mismo tamaño que se pueden tomar de la misma población.

Ello explica que, el nivel de error de los resultados con una muestra pueden ser y de hecho lo serán frecuentemente menores que el error máximo señalado por el error muestral.

FUNDAMENTO MATEMATICO-ESTADISTICO DEL ERROR MUESTRAL

El error muestral tiene el fundamento estadístico siguiente:

Llamemos $\bar{x}$ a la media de la muestra y $\bar{X}$ a la media de la población.

Supongamos que se obtuvieran todas las muestras posibles de n unidades cada una, de la misma población, y se calculasen las $\bar{x}$ para cada una de ellas. Entonces las estimaciones de $\bar{x}$ variarán en torno a $\bar{X}$ en la forma de curva de distribución normal.

Esta curva expresa gráfica y matemáticamente la distribución de todas las medias de las muestras obtenidas del universo. Representa que se obtendrían muchos casos de medias iguales o muy cercanas a la media de la población. En cambio indica que los casos de medias de las muestras relativamente alejados de la media de la población serían bastante más reducidos, hasta llegar a ser muy raros en los casos muy alejados por exceso o por defecto de la media del universo.

Supuesto esto, el error muestral no es otra cosa que la desviación típica de las distintas medias obtenidas alrededor de la media de la población, y se designa por $\sigma_{\bar{x}}$.

Mide la importancia de las variaciones de las $\bar{x}$ o medias muestrales en torno a la $\bar{X}$ media de la población.

Como dice H. M. Blalock (1966, p. 154), si consideramos las medias de las muestras como apreciaciones de la media de la población, podemos decir que hay cierta cantidad de error en nuestro proceso de estimación, debido a fluctuaciones de muestreo. Por consiguiente, designamos la desviación típica de una distribución de muestreo como error estandar.

FORMULAS DEL ERROR MUESTRAL

Las fórmulas generales son:

$$\frac{\sigma}{\sqrt{n}} \quad \text{y} \quad \sqrt{\frac{p \cdot (1-p)}{n}}$$

Sin embargo, estas fórmulas generales presentan modalidades diversas según se trate de muestras con desviación típica o muestras con proporciones. E igualmente según el universo ser infinito o finito.

	Fórmulas del error — *Poblaciones*	
Muestras	*Infinitas*	*Finitas*
De intervalo	$e = \sqrt{\frac{\sigma^2}{n}}$	$e = \sqrt{\frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}}$
De proporciones	$e = \sqrt{\frac{p \cdot q}{n}}$	$e = \sqrt{\frac{p \cdot q}{n} \cdot \frac{N-n}{N-1}}$

En las que σ^2 y $p \cdot q$, $q = 1 - p$, son las varianzas, o cuadrado de la desviación típica, N es el tamaño de la población y n el tamaño de la muestra. Los resultados obtenidos hay que multiplicarlos por el nivel de confianza elegido 2 ó 3 sigmas. (Ver ejercicios 50 a 53).

ESTUDIO DE LOS FACTORES DEL ERROR MUESTRAL

El error muestral según las fórmulas expuestas depende de los siguientes factores:

A) La amplitud del universo, finito o infinito.

B) El tamaño, n, de la muestra.

C) El nivel de confianza adoptado.

D) La desviación típica.

A) Son finitos los universos de 100.000 unidades o menos e infinitos los que exceden de esta cantidad. El hecho de que se estime el universo como infinito supone que su amplitud no influye para nada en la fórmula a aplicar, al contrario de los que ocurre en los finitos.

B) El tamaño de la muestra es, en las fórmulas del error muestral, un dato, que nos dan o elegimos. Claro que si al calcular el error, éste resulta demasiado elevado, p. ej. superior al 6 % , entonces tendremos que aumentar el tamaño, para disminuír el error.

Cuanto más elevado sea el tamaño de la muestra menos es el error muestral. Hay entre ellos una relación, inversamente proporcional.

C) El nivel de confianza adoptado. Es también otro dato que nos dan o elegimos.

Las pruebas empíricas realizadas muestran que la distribución de las medias de todas las muestras posibles del mismo tamaño en una población, se ajusta generalmente a la ley normal de probabilidad con unos valores centrales medios elevados y unos valores extremos reducidos, y adopta, por tanto, la forma de una curva de campana de Gauss. Supuesto esto, el nivel de confianza no es otra cosa que la proporción del total de las medias muestrales, que forman el área de esta curva de distribución estadística normal del universo que se piensa abarcar.

Para comprender mejor esto, examinemos con más detalle la curva normal.

En esta curva se pueden distinguir:

a) La ordenada o vertical central que la divide en dos partes iguales y que representa la media del universo o el valor central de la distribución.

Con ella debe coincidir la media de las medias de las muestras si los procedimientos de elección de la muestra no son sesgados.

b) La abscisa de la curva normal que es la recta que constituye su base o pie.

c) Las ordenadas distintas de la ordenada central y que expresan diversos valores de las medias de las muestras, obtenidos más o menos frecuentemente según su mayor o menor alejamiento de la ordenada central.

d) El área general de la curva que es el área del espacio comprendido entre la curva y la recta abcisa, base o pie de la curva, que comprende toda la población o lo que es lo mismo todos los casos de la distribución de las medias muestrales.

e) Las áreas parciales de la curva o de los espacios de ésta comprendidos entre la ordenada central o media y las demás ordenadas.

Estas áreas parciales presentan una importante propiedad. Esta consite en que laproporción del área entre la media y una ordenada determinada es constante, cualquiera que sea la forma concreta de la curva que cambia con la media y la desviación típica de la población representada. La distancia de las ordenadas a la ordenada central se mide en unidades sigma o desviaciones típicas de esta curva normal.

Es concreta, la distancia de una sigma a ambos lados de la media, abarca el 68% del área de la curva normal; la de dos sigmas el 95,50% y la de 3, el 99,7 % , y así hasta el infinito porque la curva se acerca cada vez más a la base pero nunca la llega a alcanzar del todo.

Por tanto, estadísticamente nunca se puede abarcar toda el área de la curva normal. De ahí que en las operaciones estadísticas y de modo concreto en la determinación del error y del tamaño de la muestra, sea necesario determinar el área de la misma que se pretende abarcar. Esta área elegida recibe el nombre de nivel de confianza, porque representa el % de seguridad o probabilidad que elegimos. En concreto significa que de las medias de todas las muestras posibles que forman en su conjunto la curva muestral, solo consideramos como probables el 95,5 % ó el 99,7 % , por lo que prescindimos del otro 4,5 ó el 0,3 % por estimar que es muy improbable su obtención, y por creer que el nivel indicado del 95,5 ó 99,7 % nos proporciona una seguridad no total pero suficiente en la práctica. En consecuencia, con esta decisión aceptamos un riesgo razonable de equivocarnos, pues para lograr mayor seguridad tendríamos que aumentar en proporción muy elevada el tamaño de la muestra con todos los gastos e inconvenientes que esto lleva consigo.

El nivel de confianza más corriente con que se trabaja en las muestras y en las investigaciones de Sociología es el de dos sigmas que, como acabo de indicar abarca el 95,5 % del área de la curva normal. También se emplea a veces, cuando se desea mayor seguridad, el de 3 sigmas, que abarca el 99,7 % .

D) Respecto a la desviación típica hay que distinguir las muestras referentes a variables cuantitativas con escalas por tanto de intervalo o de razón, de las que se refieren a variables cualitativas con escala nominal u ordinal, de las que son tipo las variables dicotómicas, con una escala que tienen sólo dos categorías la afirmativa y la negativa.

Estas variables se pueden en cierto modo cuantificar, transformando sus categorías, de acuerdo con la nueva álgebra de Boole, en 1 la positiva y en 0 la negativa.

En estas variables cuantificadas de este modo, la media es igual a p ó la proporción en que se da el atributo en cuestión en la población y la desviación típica es igual a $\sqrt{p \cdot (1 - p)}$.

En las muestras primeras la desviación típica es σ y su fórmula es la conocida por la estadística.

Tanto σ como $\sqrt{p \quad (1 - p)}$ se refieren a la varianza del universo a la que representan. Cuándo ésta no es conocida, como suele ocurrir, se utiliza en su lugar la llamada cuasivarianza, o sea la varianza estimada mediante una pequeña muestra tomada de la población, ya que está comprobado que este parámetro es muy estable y varía muy poco aunque se utilicen muestras pequeñas.

Por ello, en el caso de las muestras del segundo tipo, para hallar la desviación típica es necesario hallar, sino se conoce p, la proporción o

el tanto por ciento que expresa el nivel estimado de la extensión en el universo de la característica sobre la que se desea obtener información.

Por ejemplo, en una encuesta sobre la extensión del consumo de una determinada marca de cigarrillos se deberá estimar inicialmente, aunque sea de manera aproximada, mediante un sondeo previo, la proporción de fumadores y no fumadores dentro del total de la población.

Cuando no se puede realizar esta estimación previa, se adopta la suposición de que dicha proporción es de 50 % , que es el caso más desfavorable, es decir, aquel en que la muestra deberá ser mayor.

A mayor desviación típica, mayor error. Hay una relación directamente proporcional.

Error en las muestras estratificadas.— En la fórmula del error de estas muestras hay que tener en cuenta que, normalmente, el %o proporción de la característica o la desviación típica en su caso, serán distintas para cada estrato. Por ello, las anteriores fórmulas se transforman en las siguientes.

Caso de proporciones: $$\text{e. est.} = \sqrt{\frac{\Sigma\, n_i\, p_i\, (1 - p_i)}{n}}$$

en la que

n_i = elementos del estrato i
p_i = la proporción del atributo muestral en el estrato i
n = el número de elementos total de la muestra
Σ = sumatorio de todos los estratos.

caso de medias: $$\text{e. est.} = \sqrt{\frac{\Sigma\, (n_i\, \sigma_i)^2}{n^2}}$$

En la que n_i y n tienen la misma significación que antes y σ_i es igual a la desviación típica del estrato i.

Se trata en este caso de muestras estratificadas, de fórmulas aproximadas, pero pueden usarse sin reparo, pues dan una seguridad suficiente. Si se calcula el error de la muestra sin estratificar, se ve que el error es mayor, de donde se deduce que la estratificación disminuye el error muestra. (Ver ejercicios 59, 60).

Error de las muestras de clusters.— En este caso la fórmula del error es la siguiente:

$$E = \sqrt{\frac{\sigma^2}{m}\ \left(1 - \frac{m}{M}\right)}$$

m = número de clusters de la muestra
M = número de clusters del universo
σ = desviación standard de las medias de los clusters de la muestra.

El error de una muestra de este tipo es siempre superior al de una muestra simple. (Ver ejercicios 57 y 58).

ERROR SISTEMATICO

En su sentido más general, error es cualquier equivocación o el hecho de tomar una cosa por otra en nuestros actos y pensamientos, pensando y actuando desacertadamente. Los errores se pueden clasificar, según su carácter, en sistemáticos, aleatorios y muestrales.

El error sistemático, se suele ajustar a un sistema o pauta determinado en su sentido y magnitud, y, por tanto, al repetirse los casos, su magnitud aumenta acumulativamente. El error aleatorio, a diferencia del anterior, varía normalmente, en cada caso distinto, en su sentido y magnitud, y por ello tiende a anularse cuando se trata no de un sólo caso, sino de muchos. El error muestral acabado de exponer, es un error de tipo aleatorio estadístico, probable y genérico, que tiene lugar cuando se utilizan muestras y se basa en las fluctuaciones aleatorias del muestreo.

Siguiendo al Profesor Bugeda (1974, p. 32 y ss.), las principales vías de penetración del error sistemático o de distorsión son las siguientes:

Primera: Modo de elección defectuoso de una muestra que da lugar a una muestra no representativa.

Segunda: La dependencia del método de elección de una característica de las unidades del universo a elegir. Ejemplo, la elección al azar sistemático, v.g. uno por cada cinco, cuando en la base de la muestra también uno de cada cinco son individuos con una característica determinada. Resultará por tanto que estos individuos se hallarán excesivamente representados en la muestra y no lo estarán suficientemente otros individuos de atributos diferentes.

Tercera: Las tendencias subjetivas conscientes o inconscientes del investigador.

Esta fuente de distorsión obliga a realizar la elección por procedimientos lo más objetivos posibles. Cada uno tiene sus preferencias conscientes o no, que llevan a una elección viciosa de la muestra cuando ésta depende en algún modo del arbitrio del investigador.

Cuarta: Las sustituciones. Se produce cuando un entrevistador se permite sustituir según su criterio las unidades de la muestra elegidas al azar.

Quinta: La insuficiente observación del conjunto de la muestra, que lleva a atenerse a los primeros resultados.

Este motivo se refiere principalmente a las encuestas por cuestionario escrito. En ellas, se ha comprobado que los que contestan primero son los más interesados en la encuesta. Por ello, en las encuestas por correo se estima necesario reiterarlas hasta tres veces, pues de otro modo además de quedar modificado el tamaño de la muestra. la encuesta sólo abarcaría un tipo de gente, los rápidos en contestar, y quedarían fuera los no rápidos, que seguramente constituyen un grupo de características diferentes de los primeros.

Estos errores se refieren a la elección de la muestra y la ejecución. Pero no sólo son posibles y se dan de hecho errores en esta fase de la investigación social, sino en todas sus fases. A estos errores se les suele dar el nombre de errores ajenos al muestreo.

A continuación se hace una enumeración de algunas de las principales modalidades de estos errores en las distintas fases de la investigación.

Operaciones básicas fundamentales: Hipótesis inadecuadas y mal formuladas.

Elección de indicadores no debidamente representativos de las variables generales a investigar.

Imprecisiones y errores en la delimitación temporal y especial del campo de investigación.

Definición ambigua de las unidades de observación y del conjunto de la población a investigar.

Observación: Elección de técnicas de observación inadecuadas.

Defectos de construcción de los instrumentos de observación.

Preguntas redactadas mal en los cuestionarios que pueden causar confusiones en los encuestados.

Empleo en ellos de palabras que levanten prejuicios en los encuestados, y les predispongan a contestaciones inexactas.

Colocación juntas en el cuestionario de preguntas que puedan influirse mutuamente entre sí.

Influencia de los entrevistadores en el sentido de las respuestas de los entrevistados.

Clasificación y análisis: Codificación y perforación defectuosas que mezclen unos resultados con otros indebidamente.

Proceso de datos sin revisión general previa de éstos.

Defectuosa explotación de los cuestionarios.

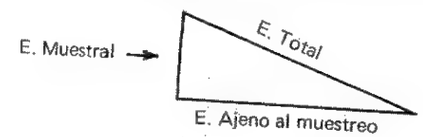
Omisión de formación de tablas de interés mientras se presentan otras menos interesantes.

Elección de procedimientos de análisis estadístico inadecuados.

Ejecución e interpretación incorrectas de los mismos.

Defectos lógicos en las deducciones y consecuencias teóricas derivadas de la investigación.

La suma del error muestral y de los errores ajenos al muestreo forma el error total de la encuesta que se suele representar por un triángulo:



EJERCICIOS

53

En una encuesta realizada respecto a una muestra de 5.056 familias ha resultado que un 45 por 100 de ellas tiene televisión. El universo era de 50.000 familias, el nivel de confianza tres sigmas y p se estimó previamente era el 55%.

Calcular el error muestral y exponer su significado. Contrastar el error obtenido con que resulte según las tablas para un tamaño de muestra similar.

RESPUESTA

Aplicando la fórmula para poblaciones finitas y % e integrando en ella el nivel de confianza que se representa por σ: $E = \sqrt{\frac{\sigma^2 \cdot p \cdot q}{n} \cdot \frac{N-n}{N-1}}$ se tiene:

$$E = \sqrt{\frac{9 \times 45 \times 55}{5.056} \times \frac{50.000 - 5.056}{50.000 - 1}} = \sqrt{4{,}40 \times 0{,}89} = \sqrt{3{,}91} = 1{,}98$$

Significación: El resultado obtenido indica que hay una probabilidad del 997 por 1.000 de que el porcentaje de las personas que tienen televisión en el universo esté comprendido en 45 ± 1,98, o sea, en el intervalo de confianza de 43,02 al 46,98 por 100.

Las tablas están hechas para poblaciones infinitas en las que no se aplica el segundo término (N − n/ N − 1), llamado de corrección. Por ello el error que corresponde a la fila de una muestra de 5.000 y una proporción de la característica del 45/55, es algo mayor 2,10.

54

En una población de 300.000 habitantes en la que p = 50, se trabaja con una muestra de 7.815 elementos al nivel de confianza de 95,5 por 100. Se interesa hallar el error muestral, aplicando la fórmula y comprobando el resultado con las tablas.

RESPUESTA

Fórmula para poblaciones infinitas y % , con el nivel de confianza 2σ integrado en ella: $E = \sqrt{\frac{\sigma^2 \cdot p \cdot q}{n}}$

$$E = \sqrt{\frac{4 \times 50 \times 50}{7.815}} = \sqrt{1{,}28} = 1{,}13$$

Utilizando la tabla para errores muestrales de dos sigmas, la intersección de la columna del número de elementos de la muestra de valor más aproximado a 7.815, que es 7.500, con la columna p/q = 50/50, da 1,2. Ambos resultados concuerdan habida cuenta de la diferencia en el tamaño de la muestra.

55

En un universo de 1.000 reclutas se ha obtenido una muestra al azar de 15. El peso de mayor a menor de esta muestra de reclutas es el siguiente, en kilogramos: 50, 54, 56, 58, 60, 61, 62, 63, 64, 65, 66, 68, 72, 74 y 78.

Hallar el error que representa la muestra y expresar su significación a un nivel de confianza de dos sigmas.

RESPUESTA

No se trata de atributos que se dan en una cierta proporción en el universo, por ejemplo, ser rubio o moreno, sino de una variable continua, el peso, que puede presentar múltiples valores. Por tanto, el error de la muestra hay que hallarlo en relación a su media, y no, como en los casos anteriores, los más frecuentes en Sociología, con relación a proporciones.

Fórmula: $E = \sqrt{\frac{\sigma^2}{n-} \cdot \frac{N-n}{N-1}}$

$$\text{Media de la muestra} = \frac{50 + 54 + 56 \ldots + 78}{15} = \frac{951}{15} = 63{,}4.$$

Media	*Pesos unidades muestra*	*Diferencia entre columnas 1 y 2*	*Diferencia al cuadrado*
63,4	50	— 13,4	179,56
63,4	54	— 9,4	88,36
63,4	56	— 7,4	54,76
63,4	58	— 5,4	29,16
63,4	60	— 3,4	11,56
63,4	61	— 2,4	5,76
63,4	62	— 1,4	1,96
63,4	63	— 0,4	0,16
63,4	64	0,6	0,36
63,4	65	1,6	2,56
63,4	66	2,6	6,76
63,4	68	4,6	21,16
63,4	72	8,6	73,96
63,4	74	10,6	112,36
63,4	78	14,6	213,16
		TOTAL	801,60

$$\sigma^2 = \frac{801{,}6}{15} = 53{,}44$$

Error muestral:

$$E = \sqrt{\frac{53{,}44}{15} \times \frac{1.000 - 15}{1.000 - 1}} = \sqrt{3{,}56 \times 0{,}98} = \sqrt{3{,}48} = 1{,}86$$

Significación a un nivel de confianza de dos sigmas: Indica que hay una probabilidad del 95,5 por 100 de que la media del universo esté comprendida entre 63,4 ± ± (1,86 x 2), o sea, entre 59,68 y 67,12 kilogramos.

56

En el universo de 150.000 habitantes mayores de edad se ha obtenido una muestra de 10.000 para conocer la renta mensual por habitante de dicho universo. Esta ha resultado de 8.000 pesetas, y la desviación típica referente a esta medida es de 1.200.

Se pide hallar el error que representa esta muestra y expresar cuantitativamente su significado a un nivel de confianza dos y tres sigmas, teniendo en cuenta que la fórmula a aplicar por tratarse de un universo considerado infinito es:

$$E = \sqrt{\frac{\sigma^2}{n}}$$

RESPUESTA

Aplicando la fórmula se tiene:

$$E = \sqrt{\frac{1.200^2}{10.000}} = \frac{1.200}{100} = 12;\ E = 12 \text{ en unidades sigma.}$$

Interpretación: A un nivel de confianza de 2 sigmas (12 x 2 = 24), indica que hay una probabilidad del 95,5 por 100 de que la media de renta del universo esté comprendida entre 8.000 ± 24, o sea, entre 7.976 y 8.024.

A un nivel de confianza de tres sigmas, 3 x 12 = 36, indica que hay una probabilidad de 997 por 1.000 de que la renta media del universo está comprendida entre 8.000 ± 36, o sea, 8.036 y 7.974.

57

Se va a hacer un estudio de una comarca asturiana que comprende 120 aldeas. Para ello se ha elegido una muestra de ocho. Las dimensiones relativas de cada poblado son las siguientes: 2,4, 3,6, 4,1, 4,3, 4,8, 5,1, 6 y 6,2.

Indicar: 1. Qué tipo de muestra se trata. 2. Hallar el error que representa.

DETERMINACION DEL MARGEN DE ERROR (*E*) DE UN RESULTADO, SEGUN EL NUMERO DE ELEMENTOS COMPRENDIDOS EN LA MUESTRA (1.ª COLUMNA) Y DE LOS VARIOS VALORES *p* (INDICADOS EN EL TITULO)

Fórmula: $E = 3\sqrt{\frac{p\,q}{N}}$

Seguridad del 99,7 % (3 sigma)

Número de elementos de la muestra	1/99	2/98	3/97	4/96	5/95	10/90	15/85	20/80	25/75	30/70	35/65	40/60	45/55	50/50
25	5,9	8,4	10,3	11,8	13,1	18,0	21,4	24,0	26,0	27,5	28,6	29,4	29,8	30,0
50	4,2	6,0	7,3	8,3	9,3	12,8	15,2	17,0	18,4	19,5	20,3	20,9	21,1	21,3
75	3,4	4,8	5,9	6,8	7,5	10,4	12,3	13,8	14,9	15,8	16,5	16,9	17,1	17,3
100	3,0	4,2	5,1	5,9	6,5	9,0	10,7	12,0	13,0	13,7	14,3	14,7	14,9	15,0
150	2,4	3,4	4,2	4,8	5,4	7,4	8,8	9,8	10,7	11,3	11,7	12,1	12,2	12,3
200	2,1	3,0	3,6	4,2	4,6	6,4	7,6	8,5	9,2	9,8	10,2	10,4	10,6	10,7
250	1,9	2,6	3,2	3,7	4,1	5,7	6,7	7,6	8,2	8,7	9,0	9,3	9,4	9,5
300	1,7	2,4	3,0	3,4	3,8	5,2	6,2	7,0	7,5	8,0	8,3	8,5	8,6	8,7
400	1,5	2,1	2,6	2,9	3,3	4,5	5,4	6,0	6,6	6,9	7,2	7,4	7,4	7,5
500	1,3	1,9	2,3	2,6	2,9	4,1	4,8	5,4	5,8	6,2	6,4	6,6	6,7	6,8
600	1,2	1,7	2,1	2,4	2,7	3,7	4,4	4,9	5,3	5,6	5,9	6,0	6,1	6,2
800	1,0	1,5	1,8	2,1	2,3	3,2	3,7	4,2	4,5	4,8	5,0	5,1	5,2	5,3
1.000	0,95	1,3	1,6	1,9	2,1	2,9	3,4	3,8	4,2	4,4	4,6	4,7	4,8	4,8
1.200	0,86	1,2	1,5	1,7	1,9	2,6	3,1	3,5	3,8	4,0	4,1	4,3	4,3	4,4
1.500	0,77	1,1	1,3	1,5	1,7	2,3	2,8	3,1	3,4	3,6	3,7	3,8	3,9	3,9
2.000	0,65	0,92	1,1	1,3	1,4	2,0	2,4	2,6	2,9	3,0	3,1	3,2	3,3	3,3
2.500	0,59	0,84	1,0	1,2	1,3	1,8	2,1	2,4	2,6	2,7	2,9	2,9	3,0	3,0
3.000	0,54	0,76	0,93	1,1	1,2	1,6	1,9	2,2	2,4	2,5	2,6	2,6	2,7	2,7
4.000	0,47	0,66	0,81	0,92	1,0	1,4	1,7	1,9	2,1	2,2	2,3	2,3	2,4	2,4
5.000	0,42	0,60	0,73	0,73	0,93	1,3	1,5	1,7	1,8	2,0	2,0	2,1	2,1	2,1
7.500	0,34	0,48	0,59	0,68	0,75	1,0	1,2	1,4	1,5	1,6	1,6	1,7	1,7	1,7
10.000	0,30	0,42	0,51	0,59	0,65	0,90	1,1	1,2	1,3	1,4	1,4	1,5	1,5	1,5
15.000	0,24	0,34	0,42	0,48	0,54	0,74	0,89	0,98	1,1	1,1	1,2	1,2	1,2	1,2
25.000	0,19	0,25	0,32	0,37	0,41	0,57	0,67	0,76	0,82	0,87	0,90	0,93	0,94	0,95
50.000	0,12	0,17	0,21	0,24	0,26	0,36	0,43	0,48	0,51	0,56	0,57	0,59	0,60	0,60

DETERMINACION DEL MARGEN DE ERROR (*E*) DE UN RESULTADO, SEGUN EL NUMERO DE ELEMENTOS COMPRENDIDOS EN LA MUESTRA (1.ª COLUMNA) Y DE LOS VARIOS VALORES *p* (INDICADOS EN EL TITULO)

Fórmula: $E = 2\sqrt{\frac{p\,q}{N}}$

Seguridad del 95,5 % (2 sigma)

Número de elementos de la muestra	1/99	2/98	3/97	4/96	5/95	10/90	15/85	20/80	25/75	30/70	35/65	40/60	45/55	50/50
25	4,0	5,6	6,8	7,8	8,7	12,0	14,3	16,0	17,3	18,3	19,1	19,6	19,8	20,0
50	2,8	4,0	4,9	5,6	6,2	8,5	10,1	11,4	12,3	13,0	13,5	13,9	14,1	14,2
75	2,3	3,2	3,9	4,5	5,0	6,9	8,2	9,2	10,0	10,5	11,0	11,3	11,4	11,5
100	2,0	2,8	3,4	3,9	4,4	6,0	7,1	8,0	8,7	9,2	9,5	9,8	9,9	10,0
150	1,6	2,3	2,8	3,2	3,6	4,9	5,9	6,6	7,1	7,5	7,8	8,0	8,1	8,2
200	1,4	2,0	2,4	2,8	3,1	4,3	5,1	5,7	6,1	6,5	6,8	7,0	7,0	7,1
250	1,2	1,8	2,2	2,5	2,7	3,8	4,5	5,0	5,5	5,8	6,0	6,2	6,2	6,3
300	1,1	1,6	2,0	2,3	2,5	3,5	4,1	4,6	5,0	5,3	5,5	5,7	5,8	5,8
400	0,99	1,4	1,7	2,0	2,2	3,0	3,6	4,0	4,3	4,6	4,8	4,9	5,0	5,0
500	0,89	1,3	1,5	1,8	2,0	2,7	3,2	3,6	3,9	4,1	4,3	4,4	4,5	4,5
600	0,81	1,1	1,4	1,6	1,8	2,5	2,9	3,3	3,6	3,8	3,9	4,0	4,1	4,1
800	0,69	0,98	1,2	1,4	1,5	2,1	2,5	2,8	3,0	3,3	3,3	3,4	3,5	3,5
1.000	0,63	0,90	1,1	1,3	1,4	1,9	2,3	2,6	2,8	2,9	3,1	3,1	3,2	3,2
1.200	0,57	0,81	0,99	1,1	1,3	1,7	2,1	2,3	2,5	2,7	2,8	2,8	2,9	2,9
1.500	0,51	0,73	0,89	1,0	1,1	1,6	1,9	2,1	2,3	2,4	2,5	2,5	2,6	2,6
2.000	0,44	0,61	0,75	0,86	0,90	1,3	1,6	1,8	1,9	2,0	2,1	2,2	2,2	2,2
2.500	0,40	0,56	0,68	0,78	0,87	1,22	1,4	1,6	1,7	1,8	1,9	2,0	2,0	2,0
3.000	0,36	0,51	0,62	0,71	0,79	1,1	1,3	1,5	1,6	1,7	1,7	1,8	1,8	1,8
4.000	0,31	0,44	0,54	0,62	0,69	0,95	1,1	1,3	1,4	1,4	1,5	1,5	1,6	1,6
5.000	0,28	0,40	0,49	0,56	0,62	0,85	1,0	1,1	1,2	1,3	1,4	1,4	1,4	1,4
7.500	0,23	0,32	0,39	0,45	0,50	0,69	0,82	0,92	1,0	1,1	1,1	1,1	1,1	1,2
10.000	0,20	0,28	0,34	0,39	0,44	0,60	0,41	0,80	0,87	0,95	0,95	0,98	0,99	1,0
15.000	0,16	0,23	0,28	0,32	0,36	0,49	0,59	0,66	0,71	0,75	0,78	0,80	0,81	0,82
25.000	0,12	0,18	0,22	0,25	0,27	0,38	0,45	0,50	0,65	0,58	0,60	0,62	0,62	0,63
50.000	0,08	0,11	0,14	0,16	0,17	0,24	0,29	0,32	0,35	0,37	0,48	0,39	0,40	0,40

RESPUESTA

La fórmula que corresponde aplicar al tipo de muestra en cuestión es la siguiente:

$$E = \sqrt{\frac{\sigma^2}{m} \cdot \left(1 - \frac{m}{M}\right)},$$

en la que m es el número de unidades de la muestra y M el del universo.

1. Se trata de muestra de "clusters", o conjuntos, en la que las unidades de la muestra no son individuos, sino aldeas.

2. Para aplicar la fórmula es necesario hallar la varianza y, por tanto, en primer lugar, la media de la muestra.

Media muestral:

$$\bar{x} = \frac{2{,}4 + 3{,}6 + 4{,}1 + 4{,}3 + 4{,}8 + 5{,}1 + 6 + 6{,}2}{8} =$$

$$= \frac{36{,}50}{8} = 4{,}56, \text{ redondeando, } 4{,}6$$

Varianza:

Unidades muestra	*Media*	*Diferencia*	*Diferencias al cuadrado*
2,4	4,6	—2,2	4,84
3,6	4,6	—1	1
4,1	4,6	—0,5	0,25
4,3	4,6	—0,3	0,09
4,8	4,6	—0,2	0,04
5,1	4,6	—0,5	0,25
6	4,6	1,4	1,96
6,2	4,6	1,6	2,56
		TOTAL	10,99

$$\sigma^2 = \frac{10{,}99}{8} = 1{,}37$$

$$\text{E. clusters} = \sqrt{\frac{1{,}37}{8} \times \left(1 - \frac{8}{120}\right)} = \sqrt{\frac{1{,}37}{8} \times \frac{112}{120}} =$$

$$= \sqrt{\frac{153,44}{960}} = \sqrt{0,16} = 0,40$$

58

En una localidad compuesta de 500 familias se ha obtenido una muestra por el método de azar simple de diez familias. El número de personas de cada una de las familias componentes de la muestra obtenida fue el siguiente: 5, 3, 2, 3, 3, 4, 3, 2, 8, 1.

Se pide: 1. Indicar qué tipo de muestra se trata. 2. Hallar el error muestral. 3. Señalar su significación. 4. Comparar el error obtenido con el que resultaría si se tratara de una muestra simple.

RESPUESTA

1. Es una muestra como la anterior de "cluster", ya que las unidades de la muestra son colectivas, las familias.

2. Error de la muestra:

Medidas	*Media*	*Diferencia*	*Diferencia al cuadrado*
5	3,4	— 1,6	2,56
3	3,4	— 0,4	0,16
2	3,4	— 1,4	1,96
3	3,4	— 0,4	0,16
3	3,4	— 0,4	0,16
4	3,4	— 0,6	0,36
3	3,4	— 0,4	0,16
2	3,4	— 1,4	1,96
8	3,4	4,6	21,16
1	3,4	— 2,4	5,76
34			34,40

Media $\overline{\chi} = 34 : 10 = 3,40$. Varianza $\sigma^2 = \frac{34,4}{10} = 3,44$.

$$\text{Error cluster} = \sqrt{(1 - \frac{10}{500}) \times \frac{3,44}{10}} = \sqrt{0,98 \times 0,34} = \sqrt{0,33} = 0,57.$$

3. Comparación con el error que resultaría de tratarse de una muestra simple. En este caso contaría en la muestra cada miembro de la familia individualmente y, por tanto, el tamaño de la muestra sería 34.

Aplicando la fórmula del ejercicio 75, se tendría:

$$E = \sqrt{\frac{3,4}{34} \times \frac{500 - 34}{500 - 1}} = \sqrt{0,10 \times 0,93} = \sqrt{0,10} = 0,33$$

El error es más reducido porque el tamaño de la muestra es mayor en este caso.

4. Significación a un nivel de seguridad de dos sigmas, hay una probabilidad del 955 por 1.000 de que la media del universo esté comprendida entre 3,4 ± 1,14, o sea, entre 2,25 y 4,54. Como se ve, se trata de un error muy elevado.

59

Dada la composición siguiente de una muestra estratificada del personal de una fábrica: técnicos, 30; empleados, 210; obreros especializados, 120, y obreros sin especializar, 340. Mediante un sondeo previo se ha estimado que leen la prensa habitualmente, un periódico por lo menos, un 90 por 100 de los técnicos, un 80 por 100 de los empleados, un 60 por 100 de los obreros especializados y un 30 por 100 de los obreros sin especializar.

Hallar el error que representa esta muestra considerándola estratificada y sin estratificar y exponer su significación.

RESPUESTA

Fórmula: $\text{E. est.} = \sqrt{\frac{\Sigma n_i \cdot p_i \cdot (1 - p_i)}{N^2}}$

Estratos	*Elementos de cada estrato,* n_i	*Proporción de cada estrato,* p_i	$n_i \times p_i$	$1-p_i$	$n_i \times p_i \times (1-p_i)$
Técnicos	30	90	2.700	10	27.000
Empleados	210	80	16.800	20	336.000
Obreros esp. ...	120	60	7.200	40	288.000
Obreros sin esp.	240	30	7.200	70	504.000
	TOTALES		33.900		1.155.000

$$\text{E. est.} = \sqrt{\frac{1.155.000}{600^2}} = \frac{1.075}{600} = 1,80$$

Error de la muestra considerada sin estratificar.

En este caso, para conocer el valor de p hay que hallar la media de los $p_i + n_i$

dividiendo este producto por 600, de donde resulta $p = \frac{33.900}{600} = 56\ \%$. Luego $q = 100 - 56 = 44$.

$$E = \sqrt{\frac{56 \times 44}{600}} = \sqrt{\frac{2.464}{600}} = \sqrt{4,1} = 2,02$$

Se ve que la utilización de la fórmula estratificada disminuye el error muestral.

En cuanto a su significación en la muestra estratificada indica, como en todos los casos considerados, en el supuesto corriente de un nivel de confianza de dos sigmas, que hay una probabilidad del 95,5 por 100 de que la media en el universo de los que leen habitualmente el periódico esté comprendida en el caso de los técnicos entre el 90 por 100 ± (1,80 x 2), o sea, entre 86,4 y 93,6; en el de los empleados entre 83,6 y 76,4 en el de los obreros especializados entre 56,4 y 63,6 y en el de los obreros sin especializar entre 26,4 y 33,6.

60

En el ejercicio 65, referente a un universo de 10.000 individuos, con cuatro estratos: uno, clase superior, 4 por 100 del total; otro, clase media, 50 por 100; el tercero, trabajadores, 40 por 100, y el cuarto, pobre, 6 por 100, que suman en total 2.000 suponiendo primero que el tanto por ciento de los que poseen lavavajillas en las distintas clases es el siguiente:

	%
Superior	80
Media	40
Trabajadora	10
Pobre	2

Y segundo, que las desviaciones típicas de cada estrato son las indicadas de 2,50 2, 3 y 3,33 y que la fórmula en este caso es:

$$E.\ est. = \sqrt{\frac{\Sigma\,(n_i\,\sigma)^2}{n^2}}$$

Se pide hallar el error de esta muestra estratificada, partiendo primero de la base de las proporciones indicadas y segundo de las desviaciones típicas.

RESPUESTA

1. Error según proporciones:

Estratos	*Elementos* n_i	*Proporciones lavavajillas* p_i	$n_i \times p_i$	$100-p_i$	$n_i \times p_i \times (1-p_i)$
Superior	80	80	6.400	20	128.000
Media	1.000	40	40.000	60	2.400.000
Trabajadora ...	800	10	8.000	90	720.000
Pobre	120	2	240	98	23.520
SUMAS	2.000		54.640		3.271.520

Sustituyendo en la fórmula que figura en el ejercicio anterior:

$$E.\ est. = \sqrt{\frac{3.271.520}{2.000^2}} = \sqrt{0,81} = 0,90 \text{ en unidades sigma.}$$

2. Error según desviaciones:

Estratos	*Elementos*	*Desviaciones*	*Columnas 2 × 3*	*Cuadrado columna 4*
Superior	80	2,5	200	40.000
Media	1.000	2	2.000	4.000.000
Trabajadora ...	800	3	2.400	5.760.000
Pobre	120	3,33	400	160.000
SUMAS	2.000			10.060.000

$$E.\ est. = \sqrt{\frac{10.060.000}{2.000^2}} = \sqrt{2,51} = 1,58$$

Este resultado no se puede comparar con el anterior, dado que tanto las proporciones como las desviaciones típicas utilizadas son supuestas y no guardan relación entre sí.

12. TAMAÑO DE LAS MUESTRAS

Una de las condiciones esenciales de la muestra es que constituya una porción del universo. Sin embargo, no basta cualquier parte del universo para formar una muestra, sino que ésta ha de constar de un número suficiente de elementos, elegidos al azar, tal que proporcione una seguridad estadística de que los resultados que se obtengan de ella, puedan, dentro de los límites estimados, representar realmente al universo.

Esto significa que el tamaño de una muestra ha de alcanzar determinadas proporciones mínimas, fijadas estadísticamente, según las leyes experimentales de la probabilidad. Pero, por otra parte, las necesidades prácticas de ahorro de tiempo, coste y esfuerzos, aconsejan que el tamaño de la muestra no exceda este límite mínimo marcado por la estadística.

El tamaño de la muestra y también su cálculo, depende de los cuatro factores o elementos siguientes, ya explicados al tratar del error muestral.

a) La amplitud del universo infinito o no.

b) El nivel de confianza adoptado.

c) El error de estimación.

d) La desviación típica.

LAS FORMULAS PARA HALLAR EL TAMAÑO

Las fórmulas para hallar el tamaño de la muestra, conjugan los factores acabados de indicar: el universo finito o infinito; el nivel de confianza; el error muestral elegido y la desviación típica.

El tamaño de las muestras también depende del tipo de muestreo que se haya elegido. Por ejemplo, en las muestras estratificadas, el error es menor que en las no estratificadas, y también lo es el tamaño de la muestra necesario.

Pero las fórmulas especiales del tamaño de la muestra varían con el tipo de muestreo y con el sistema de afijación de la muestra y resultan

complicadas. Como además en las ciencias sociales no se trata de ser técnicos en muestras, lo que constituye una especialidad en la estadística, en los manuales de técnicas de investigación social, sólo se suelen exponer las fórmulas fundamentales, que son las que se refieren a universos finitos e infinitos y un nivel de confianza de dos o tres sigmas.

En todo caso hay que tener en cuenta, que en general no hay inconveniente en aplicar las fórmulas para hallar el tamaño de las muestras en el sistema de muestreo aleatorio simple a los casos de muestreo especiales.

Con ello, por lo general, se obtendrá una muestra mayor, lo que por tanto implicará una mayor exactitud en los resultados al disminuir el error muestral.

Las fórmulas para hallar el tamaño de la muestra se deducen matemáticamente despejando n en las fórmulas del error muestral.

Son las siguientes:

Muestras	*Universos*			
	Infinitos		*Finitos*	
	Nivel de confianza		*Nivel de confianza*	
	2σ	3σ	2σ	3σ
De intervalo	$\frac{4\sigma^2}{E^2}$	$\frac{9\sigma^2}{E^2}$	$\frac{4 \cdot N \cdot \sigma^2}{4 \cdot E^2 + N\sigma^2}$	$\frac{9 \cdot N \cdot \sigma^2}{9 \cdot E^2 + N\sigma^2}$
Proporciones	$\frac{4 \cdot p \cdot q}{E^2}$	$\frac{9 \cdot p \cdot q}{E^2}$	$\frac{4 \cdot N \cdot p \cdot q}{E^2 (N-1) + 4 \cdot p \cdot q}$	$\frac{9 \cdot N \cdot p \cdot q}{E^2 (N-1) + 9 \cdot p \cdot q}$

En las que σ^2 y $p \cdot q$ son las varianzas, E, es el error, n, el tamaño de la muestra y N el tamaño de la población.

Existen también tablas que dan directamente para determinados valores de dichos elementos el tamaño de la muestra. Son distintas según se trate de universos infinitos o finitos y de un nivel de seguridad de dos sigmas o de tres, en cada caso. (Ver ejercicios 61 a 64).

AFIJACION DE LA MUESTRA

Un problema especial que plantea el tamaño de las muestras, en el caso de muestras estratificadas, es el de su "afijación", o sea, la determinación no sólo del tamaño general de la muestra que se puede establecer

según las fórmulas comunes, sino también la especificación del volumen de cada estrato de la muestra. Dado que es condición básica de la muestra que represente lo más exactamente posible al universo, en ella deben figurar los estratos según su proporción en el universo.

La forma más directa y simple de realizar esta operación consiste en aplicar el porcentaje, que representa cada estrato dentro del universo, al tamaño general de la muestra, con lo que se obtendrá el número de elementos de la muestra que se debe asignar a cada estrato.

Sin embargo, si se conocen las desviaciones típicas o las proporciones de cada estrato en el universo, o se han calculado antes al efecto, mediante un sondeo previo con relación al objeto de estudio de la muestra, otro procedimiento de afijar la muestra consiste en multiplicar el porcentaje de cada estrato en el universo por su proporción o su desviación típica y repartir el tamaño de la muestra entre los estratos proporcionalmente a dichos productos.

Este procedimiento recibe el nombre de afijación óptima y su aplicación está justificada principalmente por razones económicas, en el caso de que con su empleo se pueda reducir sensiblemente, por tener una varianza más reducida, el tamaño de los estratos que resulte más caro encuestar.

Según señala Sánchez Crespo (1976, p. 8), el sistema de afijación proporcional es el más utilizado. Además reune las siguientes ventajas: Permite una moderada ganancia global en precisión. Mantiene la igualdad de probabilidad de elección de todas las unidades del universo. (Ver ejercicio 65).

EJERCICIOS

61

Se ha proyectado realizar un estudio sociológico general de una ciudad que, según su último censo realizado, tiene 650.000 habitantes mayores de dieciséis años.

Hallar el tamaño de la muestra a obtener al efecto –teniendo en cuenta que se pretende trabajar a un nivel de confianza del 99,7 por 100 y con un margen de error permitido del 4 por 100–, mediante la aplicación de la fórmula y la utilización de las tablas y con indicación de la significación del resultado obtenido.

Interesa que se indique también el tamaño de la muestra a un mismo nivel de confianza y de error muestral si en lugar de tratarse de un universo de 650.000 habitantes tuviera éste 2.150.000.

RESPUESTA

Al ser el universo superior a 100.000, hay que utilizar la fórmula para universos infinitos siguiente:

$$N = \frac{\sigma^2 \cdot p \cdot q}{E^2}$$

Sigma, σ, vale aquí tres, al ser el nivel de confianza el 99,7 por 100.

En cuanto a p, como no se indican las proporciones que guardan dentro del universo las características a estudiar, es preciso suponer el caso más desfavorable, de p igual a 50, luego $q = 50$.

Sustituyendo en la fórmula, se tiene:

$$N = \frac{3^2 \cdot 50 \cdot 50}{4^2} = \frac{9 \cdot 2.500}{16} = 1.406{,}25,$$

redondeando 1.406.

Consultando la tabla para poblaciones infinitas con un nivel de confianza del 99 por 100 o de tres sigmas, se ve que la intersección de la fila que corresponde al error del 4 por 100 y de la columna que corresponde a la proporción 50/50 entre p y q es exactamente la cifra obtenida de 1.406.

Como en la fórmula no entra para nada la amplitud de la población, precisamente porque se consideran como infinitos todos los universos en cuanto pasen de 100.000, el tamaño de la muestra a elegir en el supuesto de que la ciudad tuviera 2.150.000 habitantes, sería exactamente el mismo que el acabado de hallar.

El resultado obtenido significa que el tamaño de la muestra debe ser por lo menos 1.406 para estar seguros con una probabilidad de 997 por 1.000 de que los resultados de la muestra son válidos, dentro de los márgenes del error admitido, para el universo.

62

Una empresa de fabricación de electrodomésticos ha encargado a una casa de investigación de mercados realizar un estudio sobre las condiciones del mercado de máquinas de afeitar eléctricas en una región en la que, según el censo, el número de varones mayores de quince años es de 200.000. Según un sondeo previo realizado, se afeitan con máquina eléctrica un 80 por 100 de dicha población.

Se pide determinar, aplicando la fórmula y con las tablas, el tamaño de la muestra a seleccionar, teniendo en cuenta que se quiere trabajar a un nivel de confianza del 95,50 por 100 y con un margen de error muestral del 3 por 100. Intérpretese también el resultado.

RESPUESTA

Se trata de un ejercicio similar al anterior, con la diferencia de que se conoce el valor de *p* o la característica a investigar, que es el 80 por 100; luego *q* será el 20. Asimismo, al ser el nivel de confianza elegida el 95,50, sigma vale dos, y no tres como antes.

Aplicando la fórmula:

$$N = \frac{\sigma^2 \cdot p \cdot q}{E^2}$$

tendremos:

$$N = \frac{4 \cdot 80 \cdot 20}{9} = \frac{6.400}{9} = 711{,}11$$

N, tamaño de la muestra, 711.

Consultando la tabla para poblaciones infinitas y un intervalo de confianza del 95,50 por 100, o dos sigmas, se ve que la intersección de la columna 20/80 y de la fila del error del 3 por 100 es exactamente 711.

El resultado obtenido significa que el tamaño de la muestra debe ser por lo menos de 711 unidades para estar seguros con una probabilidad de 955 por 1.000 de que los resultados de la muestra son válidos, dentro del margen de error admitido del 3 por 100, para el universo.

63

Se pretende realizar un estudio de las actitudes hacia la experiencia prematrimonial de los estudiantes de una Universidad española que cuenta con una población estudiantil de 10.000 alumnos.

Hallar el tamaño de la muestra, aplicando la fórmula y con las tablas, supuesto que se piensa trabajar a un nivel de confianza del 95,50 por 100 y con un margen de error permitido del 2 por 100

RESPUESTA

Al ser el universo inferior a 100.000, hay que aplicar la fórmula para poblaciones finitas siguiente:

$$n = \frac{4 \cdot p \cdot q \cdot N}{E^2 (N-1) + 4 \cdot p \cdot q}$$

E en este caso es igual a 2 y *p* y *q* igual a 50. N = 10.000.

Para población infinita o muy numerosa

EXTRACTO PARCIAL DE LA TABLA PRONTUARIO, ELABORADA POR EL HARVARD COLLEGE PARA ESTABLECER EL NUMERO DE ELEMENTOS A INCLUIR EN UNA MUESTRA, PARA ESTAR PRACTICAMENTE SEGUROS (SEGURIDAD DEL 997 por 1.000) QUE EL RESULTADO ESTE CONTENIDO DENTRO DE LOS LIMITES DE ERROR (±) INDICADOS AL MARGEN

Intervalo de confianza del 99,7 % (3 sigma) — Fórmula: $N = \frac{9\,p\,q}{E^2}$

Límites de error ± 3 σ en % (E)	VALORES PRESUMIBLES DE p Y DE q en % (p + q = 100)																	
	1/99	2/98	3/97	4/96	5/95	10/90	15/85	20/80	25/75	30/70	35/65	40/60	45/55	50/50
0,1	89.100	176.400	261.900	345.600	427.500	810.000	1.147.500	1.440.000	1.687.500	1.890.000	2.047.500	2.160.000	2.227.500	2.250.000
0,2	22.275	44.100	65.475	86.400	106.875	202.500	286.875	360.000	421.875	472.500	511.875	540.000	556.875	562.500
0,3	9.900	19.600	29.100	38.400	47.500	90.000	127.400	160.000	187.500	210.000	227.500	240.500	247.500	250.000
0,4	5.569	11.025	16.389	21.600	26.719	50.625	71.719	90.000	105.469	118.125	127.969	135.000	139.219	140.625
0,5	3.564	7.056	10.476	13.824	17.100	32.400	45.000	57.600	67.500	75.600	81.900	86.400	89.100	90.000
0,6	2.475	4.900	7.275	9.600	11.875	22.500	31.875	40.000	46.875	52.500	56.875	60.000	61.875	62.500
0,7	1.818	3.600	5.345	7.053	8.724	16.531	23.418	29.388	34.459	38.571	41.786	44.082	45.459	45.918
0,8	1.392	2.756	4.092	5.400	6.680	12.656	17.930	22.500	27.387	29.531	31.992	33.750	34.805	35.156
0,9	1.100	2.178	3.233	4.267	5.278	10.000	14.167	17.778	20.833	25.278	26.667	26.667	27.500	27.778
1,0	891	1.764	2.619	3.456	4.275	8.100	11.475	14.400	16.875	18.900	20.475	21.600	22.275	22.500
1,5	396	784	1.164	1.356	1.900	3.600	5.100	6.400	7.500	8.400	9.100	9.600	9.900	10.000
2,0	223	441	655	864	1.069	2.025	2.869	3.600	4.219	4.725	5.119	5.400	5.569	5.627
2,5	143	282	419	553	684	1.296	1.838	2.304	2.700	3.024	3.278	3.456	3.564	3.600
3,0	99	196	291	384	475	900	1.275	1.600	1.875	2.100	2.275	2.400	2.475	2.500
3,5	73	144	241	282	349	.661	937	1.176	1.378	1.543	1.671	1.763	1.818	1.837
4,0	50	110	164	216	267	506	717	900	1.055	1.181	1.280	1.350	1.392	1.406
4,5	44	87	129	171	211	400	567	711	833	933	1.011	1.067	1.100	1.111
5,0	36	71	105	138	171	324	459	576	675	756	819	864	891	900
6,0	25	49	73	96	119	225	319	400	469	525	569	600	619	625
7,0	18	36	53	71	87	165	234	294	344	386	418	441	455	459
8,0	14	28	41	54	67	127	179	225	264	295	320	338	348	352
9,0	11	22	32	43	53	100	142	178	208	233	253	267	275	278
10,0	9	18	26	35	43	81	115	144	169	189	205	216	223	225
15,0	4	8	12	15	19	36	51	64	75	84	91	96	99	100
20,0	2	4	7	9	11	20	29	36	42	47	51	54	56	56
25,0	1	3	4	6	7	13	18	23	27	30	33	35	36	36
30,0	1	2	3	4	5	9	13	16	19	21	23	24	25	25
35,0	0,7	1	2	3	3	7	9	12	14	15	17	18	18	18
40,0	0,6	1	2	2	3	5	7	9	11	12	13	14	14	14

Para población infinita o muy numerosa

TABLA PRONTUARIO ELABORADA POR TAGLIACARNE, PARA ESTABLECER SUFICIENTEMENTE SEGUROS (SEGURIDAD DEL 95,5 %) QUE EL RESULTADO ESTE COMPRENDIDO DENTRO DEL LIMITE DE ERROR (±) INDICADO AL MARGEN

Intervalo de confianza del 95,5 % (2 sigma) — Fórmula: $N = \frac{4\,p\,q}{E^2}$

Límites de error ± 2 σ en % (E)	VALORES PRESUMIBLES DE p Y DE q en % (p + q = 100)														
	1/99	2/98	3/97	4/96	5/95	10/90	15/85	20/80	25/75	30/70	35/65	40/60	45/55	50/50
0,1	39.600	78.400	116.400	153.600	190.000	360.000	510.000	640.000	750.000	840.000	910.000	960.000	990.000	1.000.000
0,2	9.900	19.600	29.100	38.400	47.500	90.000	127.500	160.000	187.500	210.000	227.500	240.000	247.500	250.000
0,3	4.400	8.711	12.933	17.067	21.111	40.000	56.667	71.111	83.333	93.333	101.111	106.667	110.000	111.111
0,4	2.475	4.900	7.275	9.600	11.875	22.500	31.875	40.000	46.875	52.500	56.875	60.000	61.875	62.500
0,5	1.584	3.136	4.656	6.144	7.600	14.400	20.400	25.600	30.000	33.600	36.400	38.400	39.600	40.000
0,6	1.100	2.178	3.233	4.267	5.278	10.000	14.167	17.778	20.833	23.333	25.278	26.667	27.500	27.778
0,7	808	1.600	2.376	3.135	3.878	7.347	10.408	13.061	15.306	17.143	18.577	19.592	20.204	20.408
0,8	619	1.225	1.819	2.400	2.969	5.625	7.969	10.000	11.719	13.125	14.219	15.000	15.469	15.625
0,9	489	968	1.437	1.896	2.346	4.444	6.296	7.901	9.259	10.370	11.235	11.852	12.222	12.346
1,0	396	784	1.164	1.536	1.900	3.600	5.100	6.400	7.500	8.400	9.100	9.600	9.900	10.000
1,5	176	348	517	683	844	1.600	2.267	2.844	3.333	3.733	4.044	4.267	4.400	4.444
2,0	99	196	291	384	475	900	1.275	1.600	1.875	2.100	2.275	2.400	2.475	2.500
2,5	63	125	186	246	304	576	816	1.024	1.200	1.344	1.456	1.536	1.584	1.600
3,0	44	87	129	171	211	400	567	711	833	933	1.011	1.067	1.100	1.111
3,5	32	64	95	125	155	294	416	522	612	686	743	784	808	816
4,0	25	49	73	96	119	225	319	400	469	525	569	600	619	625
4,5	20	39	57	76	94	178	252	316	370	415	449	474	489	494
5,0	16	31	47	61	76	144	204	256	300	336	364	384	396	400
6,0	11	22	32	43	53	100	142	178	208	233	253	267	275	278
7,0	8	16	24	31	39	73	104	131	153	171	186	196	202	204
8,0	6	12	18	24	30	56	80	100	117	131	142	150	155	156
9,0	5	10	14	19	23	44	63	79	93	104	112	119	122	123
10,0	4	8	12	15	19	36	51	64	75	83	91	96	99	100
15,0	2	3	5	7	8	16	23	28	33	37	40	43	44	45
20,0	1	2	3	4	5	9	13	16	19	21	23	24	25	25
25,0	0,6	1	2	2	3	6	8	12	12	13	15	15	16	16

TABLA PARA LA DETERMINACION DE UNA MUESTRA SACADA DE UNA POBLACION FINITA, PARA MARGENES DE ERROR DEL 1, 2, 3, 4, Y 5 POR 100, EN LA HIPOTESIS DE $p = 50\,\%$

Margen de confianza del 99,7 %

AMPLITUD DE LA POBLACION	*Amplitud de la muestra para márgenes de error abajo indicados*				
	± 1 %	± 2 %	± 3 %	± 4 %	± 5 %
500					
1.000					474
1.500				726	563
2.000				826	621
2.500				900	662
3.000			1.364	958	692
3.500			1.458	1.003	716
4.000			1.539	1.041	735
4.500			1.607	1.071	750
5.000			1.667	1.098	763
6.000		2.093	1.765	1.139	783
7.000		3.119	1.842	1.171	798
8.000		3.303	1.905	1.196	809
9.000		3.462	1.957	1.216	818
10.000		3.600	2.000	1.233	826
15.000		4.091	2.143	1.286	849
20.000		4.390	2.222	1.314	861
25.000	11.842	4.592	2.273	1.331	869
50.000	15.517	5.056	2.381	1.368	884
100.000	18.367	5.325	2.439	1.387	892
∞	22.500	5.625	2.500	1.406	900

p = proporción (en porcentaje) de los elementos portadores del carácter considerado. Si p es $< 50\,\%$, la muestra necesaria es más pequeña.

Fuente: ARKIN y COLTON, *Tables for Statisticians.*

TABLA PARA LA DETERMINACION DE UNA MUESTRA SACADA DE UNA POBLACION FINITA, PARA MARGENES DE ERROR DEL 1, 2, 3, 4, y 5 POR 100, EN LA HIPOTESIS de $p = 50\ \%$

Margen de confianza del 95,5 por 100

Amplitud de la población	*Amplitud de la muestra para márgenes de error abajo indicados*					
	± 1 %	± 2 %	± 3 %	± 4 %	± 5 %	± 10 %
500					222	83
1.000				385	286	91
1.500			638	441	316	94
2.000			714	476	333	95
2.500		1.250	769	500	345	96
3.000		1.364	811	517	353	97
3.500		1.458	843	530	359	97
4.000		1.538	870	541	364	98
4.500		1.607	891	549	367	98
5.000		1.667	909	556	370	98
6.000		1.765	938	566	375	98
7.000		1.842	949	574	378	99
8.000		1.905	976	580	381	99
9.000		1.957	989	584	383	99
10.000	5.000	2.000	1.000	588	385	99
15.000	6.000	2.143	1.034	600	390	99
20.000	6.667	2.222	1.053	606	392	100
25.000	7.143	2.273	1.064	610	394	100
50.000	8.333	2.381	1.087	617	397	100
100.000	9.091	2.439	1.099	621	398	100
∞	10.000	2.500	1.111	625	400	100

p = proporción (en porcentaje) de los elementos portadores del carácter considerado. Si p es $< 50\ \%$ la muestra necesaria es más pequeña.

NOTA.—Cuando no se indica, la cifra significa que la muestra debería tener una amplitud superior a la mitad de la población, lo que equivaldría extender la encuesta al total de la población misma.

Fuente: ARKIN y COLTON, *Tables for Statisticians.*

Sustituyendo se tiene:

$$\frac{4 \cdot 50 \cdot 50 \cdot 10.000}{4\,(10.000\text{-}1) + 4 \cdot 50 \cdot 50} = \frac{100.000.000}{49.996} = 2.000, \text{ aproximadamente.}$$

Por medio de las tablas para 2 σ.

La intersección de la fila de amplitud de la población 10.000 con la columna correspondiente al error del 2 por 100 elegida da como tamaño de la muestra el obtenido con la fórmula, 2.000.

64

a) El primer curso de la Facultad de Políticas y Sociología tiene 1.000 miembros, de los cuales 500 se orientan a Sociología y el resto a Políticas.

Hallar cuantos elementos debe tener la muestra en una investigación sobre las aspiraciones profesionales de dicho curso, al nivel de confianza del 95 por 100, o dos sigmas, y con un error permitido del 4 por 100.

Aplicar primero la fórmula y utilizar después las tablas.

RESPUESTA

Sustituyendo en la fórmula anterior para poblaciones finitas, los valores dados de sigma. 2, error muestral, 4, p 50, luego q = 50 y N 1.000, tenemos:

$$N = \frac{4 \cdot 50 \cdot 50 \cdot 1.000}{16.999 + 50 \cdot 50 \cdot 4} = \frac{10.000.000}{25.996} = 384{,}67,$$

redondeando 385.

En las tablas la intersección de la fila de universo 1.000 y la columna del error es también 385.

b) Repetir el mismo ejercicio anterior, supuesto que se quiere trabajar a un nivel de confianza del 99,7 por 100 o de tres sigmas, y con un error permitido del 5 por 100. Interpretar el resultado en comparación con el anterior.

RESPUESTA

La fórmula a aplicar en este caso es la siguiente:

$$n = \frac{9 \cdot p \cdot q \cdot N}{E^2 \cdot (N - 1) + 9 \cdot p \cdot q}$$

Sustituyendo en la fórmula se tiene:

$$n = \frac{9 \times 50 \times 50 \times 1.000}{5^2 \times (1.000 - 1) + 9 \times 50 \times 50} = \frac{22.500.000}{47.475} = 473{,}93,$$

redondeando 474.

Por medio de las tablas para tres sigmas:

La intersección de la fila 1.000 con la fórmula 5 por 100 de error da 474.

Interpretación: A pesar de que en este caso se ha elevado el error del 4 al 5 por 100, la mayor probabilidad de validez de los resultados de la muestra para el universo exigida en este caso, 997 por 1.000, frente a 955 por 1.000 en el caso anterior dentro de los márgenes de error admitidos, hace que la muestra sea sensiblemente más amplia ahora, 474, frente a 385 en el ejercicio precedente.

65

Una empresa de mercados ha de realizar, por encargo de una fábrica de aparatos lavavajillas, una encuesta en una capital de provincia en la que figuran empadronadas 10.000 familias, empleando para ello una muestra estratificada, según las siguientes categorías y proporciones. Clase superior, 4 por 100; clase media, 50 por 100; clase trabajadora, 40 por 100, y clase pobre, 6 por 100. Se pide:

1. Determinar la base y la unidad de la muestra.

2. Determinar su tamaño general por la fórmula y las tablas a un nivel de confianza del 95,5 por 100 y con un error permitido del 2 por 100.

3. Determinar la composicion de los estratos de la muestra primero de manera que sean proporcionales a los estratos del universo y después, teniendo en cuenta que las desviaciones típicas, dentro de cada estrato, respecto a la adquisición de lavavajillas es de 2,5 para el primer estrato, de 2 para el segundo, de 3 para el tercero y de 3,33 para el cuarto.

4. Determinar la forma de elegir la muestra: a) por el procedimiento de azar simple; y b) por el de cuotas.

RESPUESTA

1. La base de la muestra sería de padrón de la capital formado por las hojas familiares del censo que se realiza cada cinco años. La unidad de la muestra es, según el enunciado, la familia.

2. Para hallar el tamaño general de la muestra se aplicará la fórmula conocida suponiendo, pues no se dice nada al efecto, que la proporción de los que tienen y no tienen lavavajillas es del 50 por 100.

$$n = \frac{4 \cdot p \cdot q \cdot N}{E^2(N-1) + p \cdot q \cdot 4} = \frac{4 \cdot 50 \cdot 50 \cdot 10.000}{4 \cdot 9.999 + 50 \cdot 50 \cdot 4}$$

$$= \frac{100.000.000}{49.996}$$

n = 2.000.

Por las tablas, la intersección de la fila de la población 10.000 con la columna del error 2 por 100 es también 2.000.

3. Composición de la muestra por estratos:

$$\text{Fracción de muestreo} = \frac{2.000}{10.000} = 20\ \%$$

a) Composición proporcional:

Universo	*Estratos (%)*	*Su amplitud en el universo*	*Fracción de muestreo (%)*	*Amplitud de los estratos en la muestra*
10.000	4	400	20	80
10.000	50	5.000	20	1.000
10.000	40	4.000	20	800
10.000	6	600	20	120
		TOTAL MUESTRA		2.000

Composición no proporcional: En este caso hay que multiplicar los estratos la muestra por las desviaciones típicas respectivas, y luego repartir proporcionalmente la cifra general de la muestra a los resultados obtenidos en la multiplicación indicada:

Muestra	*% del estrato*	*Su desviación típica*	*Columna 2 por 3*	*Muestra estratificada no proporcional*
2.000	4	2,50	10	80
2.000	50	2	100	800
2.000	40	3	120	960
2.000	6	3,33	20	160
		TOTALES	250	2.000

3. Si existe clasificación del censo por estratos, se podría aplicar, como procedimiento de muestreo, bien el azar simple o el azar sistemático. Si no hay clasificación del censo por estratos ni es viable realizarla, habría que practicar el sistema de cuotas.

4. Para elección de la muestra por el sistema de azar simple se emplearía en la forma acostumbrada las tablas de números aleatorios dentro de cada estrato.

En el muestreo por cuotas se señalarían a cada entrevistador un número de entrevistas a realizar proporcional a cada estrato:

Por ejemplo:

Entrevistador: D. ...

Número de entrevistas a realizar: 48.

Estratos	*Número*
Superior	2
Media	20
Trabajadora	22
Pobre	4
TOTAL	48

Sección 4

OBSERVACION

3. Si existe clasificación del censo por estratos, se podría aplicar, como procedimiento de muestreo, bien el azar simple o el azar sistemático. Si no hay clasificación del censo por estratos ni es viable realizarla, habría que practicar el sistema de cuotas.

4. Para elección de la muestra por el sistema de azar simple se emplearía en la forma acostumbrada las tablas de números aleatorios dentro de cada estrato.

En el muestreo por cuotas se señalarían a cada entrevistador un número de entrevistas a realizar proporcional a cada estrato:

Por ejemplo:

Entrevistador: D. ...

Número de entrevistas a realizar: 48.

Estratos	*Número*
Superior	2
Media	20
Trabajadora	22
Pobre	4
TOTAL	48

Sección 4

OBSERVACION

13. OBSERVACION, MEDIDA Y ESCALAS

NOCION GENERAL DE OBSERVACION Y SUS TIPOS

Hecho el plan de la investigación a desarrollar, y resueltos los problemas que plantea el muestreo en el caso de que no se puede abarcar a todo el universo, empieza el contacto directo con la realidad objeto de investigación o trabajo de campo. Tiene por finalidad la recogida de datos mediante los diversos métodos de observación.

La observación, pues, en este sentido amplio, engloba todos los procedimientos utilizados en las ciencias sociales, no sólo para examinar las fuentes donde se encuentran los hechos y datos objeto de estudio, sino también para obtener y registrar éstos. Considerada de este modo, la observación proporciona al investigador la materia de trabajo que ha de ser objeto después de tratamiento y estudio científicos, mediante las siguientes fases de la investigación, la clasificación y tabulación y el análisis y explicación.

La realización de la observación implica, en todo caso, los siguientes elementos: un sujeto, un «objeto», el acto de la observación y el registro de lo observado. Generalmente implica también la actuación natural o provocada de un estímulo con la consiguiente respuesta del objeto observado, que se percibe y registra, dando lugar al siguiente esquema:

En todos los eslabones de esta cadena pueden intervenir factores ajenos y alterar la pureza de la observación y la validez de sus resultados. Además, frecuentemente el objeto está latente, oculto, y es observado por medio de expresiones externas que se cree lo representan. En estas circunstancias se basan las dificultades y problemas que plantea la observación, estudiadas en el capítulo 7 y la necesidad de su control.

Entendida así, la observación se puede clasificar, según su objeto, en los tipos siguientes, que abarcan los procedimientos fundamentales actualmente existentes de recogida de datos, y que serán objeto de atención sucesivamente en esta obra.

a) Observación directa por medio de los sentidos de hechos y realidades sociales presentes. Si estos hechos se producen espontánea y naturalmente, se tiene la observación simple. Pero si, por el contrario, su efectividad es provocada artificialmente por el observador, o por lo menos existe una cierta preparación y control de éste sobre las condiciones del fenómeno observado, se trata entonces de la observación experimental.

b) Observación documental, sobre documentos actuales o históricos de todo género en cuanto recogen y reflejan hechos y datos de interés social.

c) Observación, mediante encuesta, entendida esta palabra en el sentido de interrogación a los sujetos o protagonistas de los hechos estudiados. Puede tener lugar mediante cuestionario, escrito, entrevista o escala sociométrica.

Esta observación puede realizarse mediante cuestionario dirigido a amplios sectores de la población que éstos contestan por escrito. También puede efectuarse mediante entrevista, en la que los encuestados contestan de palabra. Finalmente, puede tener lugar mediante la aplicación a los encuestados, por cuestionario escrito o entrevista indistintamente, de las llamadas escalas de actitudes.

En síntesis, se puede formar el siguiente cuadro de tipos principales de observación:

Observación directa	Simple Experimental
Observación documental	
Observación mediante encuesta	Por cuestionario Por entrevista Por escala de actitudes

OBSERVACION Y MEDIDA

A primera vista puede parecer que la observación y la medida son términos distintos que no guardan una estrecha relación entre sí. Sin embargo, si se tiene en cuenta que los datos, como indicamos en el capítu-

lo segundo de esta obra, se refieren a una dimensión de una variable en una unidad de observación determinada (por ejemplo, un dato es: el encuestado X declara que pesa entre 65 y 70 Kg) y que, por tanto, los datos son la expresión verbal o escrita de medidas realizadas, se revela entonces la íntima vinculación entre la observación y la medida.

Resulta que si, conforme a la definición expuesta, la observación científica consiste en la recogida de datos, y si los datos no son otra cosa que la expresión de medidas, hay que concluir que, según lo anterior, al ser la observación una recogida de datos, se puede decir que es también una recogida de medidas.

Hasta no hace mucho, predominaba en las ciencias sociales una noción restringida tanto de la observación como de la medida. Por ejemplo, en el término dedicado a la medida en el Diccionario de Ciencias Sociales (1976, p. 187, v.2) se recoge una noción de medida estricta, según la cual "medir una cantidad es compararla con otra de igual naturaleza con objeto de averiguar cuantas veces la primera contiene a la segunda".

Igualmente la observación se suele limitar a la recogida de datos realizada mediante los sentidos humanos, es decir, a lo que hemos llamado en la anterior clasificación, observación directa.

No obstante lo anterior, no se ve razón alguna para no considerar observación toda recogida de datos, aunque no se funde directa e inmediatamente en los sentidos del investigador.

En cuanto a la medida, la definición que se ha impuesto actualmente y, sin duda, más aceptable científicamente, es la recogida por S. S. Steven (1976, p. 17), según la cual "medir es asignar números a objetos o acontecimientos de acuerdo con ciertas reglas".

Conforme a esta definición no sólo medimos cuando comparamos algo con una unidad de medida y vemos las veces que la contiene, sino también cuando distinguimos y clasificamos cosas distintas, y cuando ordenamos según más y menos las características o variables de las cosas, asignando números a las distintas cosas y variables clasificadas y ordenadas.

De ello se concluye que, científicamente, hay que admitir que existen no solo uno sino diversos niveles de medida, como veremos después, y, por tanto, diversos niveles de observación, según que en la toma de datos nos limitemos a distinguir y clasificar los seres reales, según que, además, ordenemos sus características de más a menos, o, en fin, según que no solo distingamos y midamos sino que obtengamos dimensiones mediante su comparación con una unidad de medida.

Por otra parte, lingüísticamente (Baird) la medida es una afirmación resultado de una observación y, por tanto, tiene carácter humano y falible. Por ello, no se debe decir la longitud de esta mesa es de 1,6 m., sino se ha medido esta mesa y se ha obtenido un resultado de 1,6 m. A este respecto, hay que distinguir entre contar, que permite una expresión exacta y la medida que es falible.

En conclusión, se puede afirmar que existe un vínculo íntimo entre observación y medida. Cuando al observar con los sentidos, distinguimos cosas y las clasificamos, ya estamos midiendo; cuando medimos con la mayor precisión, seguimos observando. La diferencia entre estos modos de medir no está en la esencia de la medida sino en el nivel de precisión logrado.

MATEMATICAS Y REALIDAD

Desde un punto de vista matemático, la medida representa la expresión en términos numéricos de los resultados de la observación de la realidad. Por tanto, la medida hace de enlace entre la realidad y las matemáticas.

La medida puede cumplir esta función porque hay o se puede establecer una correspondencia entre las matemáticas y la realidad.

Como escribe Stevens (1976, 18) "se puede medir solo porque, en primer lugar, hay una especie de isomorfismo entre 1) las relaciones empíricas entre los objetos y los acontecimientos y 2) las propiedades del juego formal en el que los numerales son los peones y los operadores las jugadas. Cuando esta correspondencia entre el modelo formal y su contrapartida empírica es cercana e intensa, es posible descubrir verdades sobre cosas de hecho mediante el examen del modelo mismo. Así calculamos la trayectoria de una bala o el curso de un cometa sin tocar ninguno de los dos. Y quedamos admirados del prodigioso poder de las matemáticas para ver lo que está fuera de nuestra propia visión".

En la correspondencia en cuestión entre matemáticas y realidad se pueden distinguir dos manifestaciones básicas. Los seres de la realidad y sus características o variables tienen una individualidad y sustantividad mayor o menor, real, y los números la poseen también aunque solo sea formal. Luego podemos expresar esta individualidad de los seres de la realidad por números y símbolos matemáticos.

En segundo lugar, así como las matemáticas estudian las relaciones entre números, muy diversas hoy dado el gran desarrollo actual de las ciencias exactas, así también la ciencia pretende descubrir y estudiar

las relaciones entre las características, variables, de las cosas. En consecuencia, se pueden aplicar las relaciones matemáticas entre números para representar y estudiar las relaciones entre variables de la realidad.

La adecuación de las relaciones matemáticas a las relaciones reales nunca es perfecta, pero puede alcanzar una gran precisión en las ciencias físicas, de tal modo que en ellas el simple análisis y manipulación de las relaciones matemáticas puede servir para descubrir con suficiente aproximación las propiedades y relaciones de las variables reales, y su comportamiento.

En consecuencia, parece, pues, clara la relación entre matemáticas y realidad. La conclusión que se puede deducir, por tanto, es la importancia fundamental de las matemáticas como instrumento de las ciencias modernas para el conocimiento y estudio de la realidad.

MEDIDA Y ESCALAS

En primer lugar, es preciso señalar que las medidas se refieren a las propiedades o características de las cosas observadas, es decir a las variables, que, como dijimos, son con las que trabaja la ciencia, en cuanto ésta no centra su estudio en las cosas o seres globalmente, sino que lo hace a través de sus aspectos o características, variables.

Supuesto lo anterior, se puede afirmar que así como la medida es el resultado de la observación, las escalas son el resultado de la distinción que implica la medida. En efecto, si medimos una característica o variable de diversos seres reales y anotamos los distintos resultados de esta medida, el conjunto que forman los distintos valores o categorías obtenidos forman una escala, la escala de la variable en cuestión en el caso concreto de que se trate.

Esto explica que, según vimos, una de las más importantes clasificaciones de las variables es la que se hace según la escala que forman.

En la medida como en las escalas se puede distinguir un aspecto real, empírico, y un aspecto formal, matemático. El aspecto real de la medida es la comparación en su más amplio sentido y el formal la expresión en números de los resultados de la medición.

Correlativamente el aspecto real de las escalas, son las características empíricas distinguidas y el aspecto formal los números o signos que se asignen a cada una de ellas según una regla determinada. Por ello, los distintos niveles de medida que suponen, según la definición de Steven citada, distintas reglas de asignar números a cosas, dan lugar a distintos tipos de escalas que vamos a examinar a continuación.

DISTINCION, NIVELES DE MEDIDA Y CLASES DE ESCALAS

En armonía con lo expuesto, una escala se puede definir, en sentido genérico, como la disposición conjunta de cosas distintas, pero comunes en algún aspecto. Según esta definición, las escalas son el resultado de la distinción. A su vez, se puede decir igualmente que la distinción es el resultado de la medida, tomada esta palabra en el sentido amplio del resultado de la comparación de cosas con un patrón determinado.

Por ello, así como hay diversos niveles de medida también hay diferentes tipos de distinción a los que lógicamente corresponden diversos tipos de escalas en cuanto éstas son el resultado de la distinción.

En primer lugar, existe una distinción simple, que establece únicamente la diversidad de cosas diferentes, pero iguales en algún aspecto aunque sólo sea en el más genérico de seres. Esta distinción es el resultado de la medida, en el sentido más amplio o genérico, en cuanto comparación de las cosas entre sí para establecer su equivalencia o no equivalencia, o mejor dicho, su igualdad o desigualdad. Esta distinción y medida da lugar a la enumeración conjunta de seres, objetos, cosas, especies, enumeración conjunta o escala, que se llama puramente nominal, pues el orden en que aparezca enumerado el conjunto e incluso la asignación de números de sus elementos, a efectos de clasificación, no implica de por sí ninguna significación numérica o cuantitativa.

Por ejemplo, si distinguiéramos un conjunto de partidos o tendencias políticas principales, comparándolas o midiéndolas entre sí para ver si eran iguales, y dispusiéramos los resultados obtenidos en una enumeración conjunta, obtendríamos, por ejemplo, la siguiente escala:

1. Socialismo. 2. Democracia cristiana. 3. Tradicionalismo. 4. Anarquismo 5. Comunismo.

En este caso la regla de asignación de números a cosas, a estas categorías, es simplemente la de que cada categoría tenga su número, en el ejemplo, socialismo el 1 etc., pudiendo ser éste cualquiera de los números reales existentes.

Tal escala sería de tipo nominal en la que las categorías y la asignación de números a ellas sólo supondría la distinción, sin implicar ninguna relación de orden, distancia o proporción entre ellas. Esta escala es el resultado de la operación científica más simple, la clasificación de unidades de observación en categorías distintas.

En segundo lugar, existe una distinción que, además de diferenciar, como en el caso anterior, categorías distintas, establece un orden entre ellas. Esta distinción es el resultado de la medida o determinación en dichas categorías, del mayor o menor grado en que poseen una dimensión común dada. Por ejemplo, si en la anterior escala nominal de partidos políticos, comparásemos entre sí a éstos en relación a la dimensión de mayor o menor conservadurismo, obtendríamos no ya una escala nominal en la que el orden de enumeración no tiene ninguna significación, de tal modo que se podría variar sin ninguna consecuencia, sino una escala ordinal que indica de menos a más, o de más a menos, el mayor o menor grado de conservadurismo que dichos partidos se considera poseen. En esta escala los números que se asignan a las categorías implican un orden, pero únicamente en sentido de más y menos, sin que este orden, aunque esté relacionado con números, signifique con precisión en qué cuantía es más o menos elevado el grado de conservadurismo en una categoría que en otra.

En este caso la escala sólo puede presentar una disposición inalterable, correspondiente al orden en cuestión, que en el ejemplo citado sería:

Anarquismo	1
Comunismo	2
Socialismo	3
Democracia cristiana	4
Tradicionalismo	5

La regla de la escala en este caso es asignar a cada partido un número cualquiera con la única condición de que el número de los tradicionalistas sea el más alto y así sucesivamente.

Aquí no solo se clasifican las unidades de observación en categorías distintas, sino que se ordenan estas categorías (Blalock, 1966, p. 24).

ESCALAS DE INTERVALO

En tercer lugar, hay una distinción que no sólo diferencia categorías y establece un orden entre ellas, sino también determina la distancia entre las diversas categorías. Esta distinción es el resultado de la comparación o medida de cada categoría con unidades iguales, determinadas en relación a un origen arbitrario, no natural. Este es el caso de la temperatura, cuyo origen 0 grados se determina arbitraria o convencionalmente en el punto de solidificación del agua en la escala centígrada,

mientras que no es así en la Farenheit. La escala a que da lugar esta distinción y medida es la llamada de intervalo. En las escalas de este tipo se puede determinar que la diferencia que separa dos categorías es igual o distinta a la que separa a otras.

En estas escalas dos objetos que alcancen las medidas 5 y 10 están igualmente separados o a la misma distancia que otros que alcancen las medidas 10 y 15.

Sin embargo, estas escalas no permiten establecer que la medida de una unidad de observación es doble, triple, la mitad, etc. que la medida de otra.

Para ver esto más claramente (Cfr. Nuria Cortada, 1966, p. 14) volvamos a las escalas para medir la temperatura, centígrada y Farenheit.

La correspondencia entre ambas escalas es la siguiente:

Centígrada	0	10	30	100
Farenheit	32	50	86	212

Tomemos dos puntos correspondientes de ambas escalas, 10 y 30 en la centígrada, y 50 y 86 en la Farenheit.

Dentro de cada escala el intervalo o distancia entre dichos puntos es equivalente:

$$\text{Centígrada} = \frac{30 - 10}{10 - 0} = 2; \quad \text{Farenheit} = \frac{86 - 50}{50 - 32} = 2$$

En cambio la razón o proporción entre dichos puntos o medidas no es la misma. Para la escala C. es 30 : 10 = 3. Para la escala Farenheit es 86 : 50 = 1,70.

Es decir existe una distancia constante pero no una proporción constante para los dos escalas.

ESCALAS DE RAZON O PROPORCION

En cuarto lugar, se da una distinción, que además de diferenciar categorías y establecer un orden y distancia entre ellas, permite su comparación proporcional, o sea, determinar en qué proporción es mayor una

categoría de la escala que otra. Esta distinción es el resultado de la comparación o medida en sentido estricto, pues es la que reúne todas las propiedades de los números naturales, distinción, orden, distancia y origen, y no sólo algunas de ellas, como las anteriores.

En ella hay unidades que corresponde a intervalos iguales y tienen un origen cero natural, no arbitrario, y por tanto inalterable, a diferencia del tipo acabado de exponer, en el que, como se ha indicado, este origen es arbitrario, lo que hace pueda ser variado también convencionalmente y sustituido por otro.

La escala que corresponde a este tipo de distinción o de medida es la llamada de razón o proporción, porque permiten relacionar las distancias en términos de la proporción numérica en que cada una de ellas está contenida en las demás, mientras que en las de intervalo el hecho de que la distancia entre dos categorías esté en relación de 1 a 2, no permite decir que en la segunda se halle una dimensión doble que la primera, ya que estas distancias representan intervalos medidos en relación a un origen convencional y no natural.

Como ejemplos de escalas de razón o proporción tomemos dos escalas distintas de peso: en Kgs. y en libras.

La correspondencia en ellas entre algunos de sus valores sería la siguiente:

Kg.	0	1	2	3	4	5
Lib.	0	2,2	4,4	6,6	8,8	11

Hallamos que existe la misma relación no solo entre los intervalos o diferencias, es decir:

$$\frac{3-2}{6{,}6-4{,}4} = 0{,}454 \qquad \frac{5-4}{11-8{,}8} = 0{,}454$$

sino también entre dos puntos, o sea:

$$3 : 2 = 1{,}5 \text{ y } 6{,}6 : 4{,}4 = 1{,}50$$

Esto ocurre porque los números asociados a las escalas de razón son números verdaderos con un cero auténtico. Solamente es arbitraria la unidad de medida.

En resumen, existen cuatro niveles de medida que dan lugar a cuatro tipos de escalas: la nominal, la ordinal, la de intervalo y la de razón. (Ver ejercicio 66 a 68)

EJERCICIOS

66

Dadas las series de categorías que se indican a continuación:

a) 1, obrero; 2, no obrero.
b) 1, catolicismo; 2, judaísmo; 4, islamismo; 5, anglicanismo.
c) 1, clase alta; 2, clase media-alta; 3, clase media; 4, clase media-baja; 5, clase media-baja; 5, clase baja.
d) 1, árbol; 2, mosca.
e) 1, hierba; 2, pájaro; 3, andar; 4, estar sentado.
f) La media del cociente de inteligencia de cinco niños ha dado las siguientes puntuaciones: Niño A, 92; Niño B, 85; Niño C, 104; Niño D, 76; Niño E, 98.
g) En un cuestionario la pregunta ¿existen huelgas en esta localidad? se ha categorizado como sigue:
0. Nunca.
1. Raras veces.
2. De vez en cuando.
3. Frecuentemente.
h) En otro cuestionario figura la siguiente pregunta: ¿Nos puede decir, por favor, cuántos hijos tiene?

Ninguno.	*Cinco.*
Uno.	*Seis.*
Dos.	*Siete.*
Tres.	*Ocho y más.*
Cuatro.	

Se pide indicar razonadamente si forman en conjunto una escala, indicando, en su caso, en relación a qué variable y el tipo de escala.

RESPUESTA

a) Si es una escala en relación a la variable obrero, de tipo dicotómico, algonada, en cuanto sólo tiene dos categorías entre sí.

b) Si es una escala, respecto a la variable religión, de tipo nominal, en la que no existe orden, distancia ni origen.

c) También se trata de una escala referente a la variable clase social. En ella hay orden, pero no distancia ni origen; por ello es una escala ordinal simple.

d) Se puede considerar como una escala respecto a la variable ser vivo, bien de carácter nominal, bien de carácter ordinal simple, si se considera a las categorías según su grado en relación al orden de los seres vivos.

e) No es en realidad una escala, pues se trata de categorías en relación a dos variables ser vivo y acciones humanas, y no en relación a una sola, como sería necesario para formar una escala.

f) El cociente de inteligencia es, como su nombre indica el cociente entre la edad mental y la edad cronológica multiplicado por 100. Por tanto, como tal cociente, se puede considerar como escala de razón, con distinción, orden, distancia y origen, en cuanto uno que tenga un C. I. de 100 se puede decir que tiene el doble de C. I. que otro que tenga 50. Sin embargo, esto de ningún modo significa que se pueda afirmar que el primero tenga el doble de inteligencia que el segundo, porque la medida de la edad mental no es directa ni se basa en la existencia de unidades de inteligencia, sino que se determina indirectamente mediante la aplicación de tests.

g) También es una escala referente a la variable "existencia de huelgas". En ella hay un orden y un origen cero natural, nunca; por ello es preciso concluir, se trata de una escala ordinal con origen natural.

h) Es también una escala de la variable "Número de hijos tenidos". Como se ve, la escala se corresponde exactamente con la serie de números naturales, por lo que al poseer orden, distancia y origen natural, es preciso concluir, se trata de una escala de razón o proporción, pero discreta, no continua.

67

En las escalas que se relacionan a continuación indicar razonadamente si son nominales, ordinales o de razón, teniendo en cuenta los distintos usos a que están destinadas.

1ª Grupos de edad

De 15 hasta 20.
De 20 hasta 25.
De 25 hasta 30.
De 30 hasta 35.
De 35 hasta 40.

Uso a) Simplemente como indicador de la edad.

Uso b) En relación al grado de desarrollo de la personalidad.

RESPUESTA

Uso a) Es una escala de razón porque en ella hay orden, distancia y origen cero natural, la edad cero. En esta escala se puede decir que un individuo con treinta años tiene doble edad que otro con quince. Hay en ella razón o proporción.

Uso b) Respecto a este uso la escala es ordinal simple. Supone un orden en cuanto normalmente al desarrollo de la personalidad es proporcional a la edad. Sin embargo, no hay distancia, pues los intervalos no son iguales. No hay base para decir que un año suponga igual desarrollo de la personalidad que otro. Tampoco hay origen, ni por tanto proporción. Treinta años no suponen doble desarrollo de la personalidad que quince.

2a Profesión

Obrero.
Agricultor.
Profesión liberal.
Funcionario.
Empresario.

Uso a) Como expresión de la profesión simplemente.

Uso b) Como indicador del "status" socioeconómico.

RESPUESTA

Uso a) Es cuanto a este uso es una escala nominal, sin orden, distancia, ni origen, pues como simples profesiones son categorías que no suponen orden alguno.

Uso b) Respecto a este grupo es una escala ordinal, pues cada categoría tiene y se le puede asignar un orden en la jerarquía del estatus.

3ª Número de elecciones en el sociograma de un grupo:

2 individuos 8
2 individuos 6
3 individuos 5
6 individuos 3
8 individuos 1
3 individuos 0

Uso a) Como tal número de elecciones.

Uso b) Como índice de aceptación en el grupo.

RESPUESTA

En el uso a) la escala es de tipo de razón. En ella hay orden, distancia y origen, y se puede decir que el que recibe ocho elecciones tiene doble número de elecciones que el elegido cuatro veces.

Como índice de aceptación en el grupo es de tipo ordinal, pues no tiene distancia ni origen, ya que no se puede decir que la diferencia de liderazgo de ocho a seis sea la misma que la de cinco a tres, ni que el que los que reciben seis elecciones ostenten doble liderazgo que los que reciben tres.

68

Dadas las escalas que se indican a continuación, se pide indicar simplemente su clase y propiedades.

a) La formada por un equipo de ciclistas por el número dorsal asignado a cada corredor 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

b) La formada por las siguientes categorías, que se han distinguido en un estudio sociológico sobre enfermedades mentales, y clase de ocupación.

1. *Esquizofrenia.*
2. *Maniaco depresivo.*
3. *Paranoia.*
4. *Demencia senil.*

c) La resultante de la medición de la talla de los reclutas llegados a un cuartel: 1.832 m., 1.725, 1.680, 1.640, etc.

d) La resultante de la aplicación a un grupo de sujetos de la escala arbitraria de puntos de Watson para medir la actitud político-moral:

	Puntos
Sujeto A	*38*
Sujeto B	*42*
Sujeto C	*34*
Sujeto D	*26*

e) La derivada de la aplicación a un grupo de estudiantes de la escala de distancia social de Bogardus respecto a los "hippies".

Número de estudiantes	*Puntuación*
5	*1*
6	*2*
9	*3*
15	*4*
25	*5*
12	*6*
35	*7*

RESPUESTA

a) Es una escala nominal, en la que los números no designan orden, distancia, ni origen, pues consiste en la asignación arbitraria de un número a cada corredor a efectos de distinción.

b) Es también una escala nominal, que implica únicamente distinción de categorías, pero sin establecer orden entre ellas, ni distancia, ni origen.

c) La talla es una escala de razón, en la que existe orden, distancia o intervalos iguales, centímetros, y origen cero natural, el cero de estatura.

d) y e) Se trata de escalas ordinales. En general todas las escalas de actitudes, son en realidad ordinales, aunque algunas escalas experimentales como las de tipo Thurstone se suelen considerar, por su construcción, como de intervalo. Sólo permiten establecer un orden en los individuos a que se aplican, pero sin que exista distancia o intervalos iguales en ellas, pues no se puede decir que la diferencia de actitud que separa la puntuación 38 de la 42 sea igual a la que existe entre 36 y 32, por ejemplo.

14. OBSERVACION DIRECTA SIMPLE

NOCION Y RASGOS

La observación simple es la inspección y estudio realizado por el investigador, mediante el empleo de sus propios sentidos, especialmente de la vista, con o sin ayuda de aparatos técnicos, de las cosas y hechos de interés social, tal como son o tienen lugar espontáneamente, en el tiempo en que acaecen y con arreglo a las exigencias de la investigación científica.

La observación simple es el tipo de observación más puro de la investigación social, de ahí su importancia e interés.

No obstante lo anterior, la observación presenta los siguientes inconvenientes en la investigación social:

- La limitación de los sentidos humanos respecto a la extraordinaria amplitud de los fenómenos sociales y sus múltiples facetas.
- La dificultad de captar en profundidades las relaciones sociales.
- Los frecuentes errores de percepción anejos a ella y difícilmente evitables en la investigación social.
- Por último, el hecho de que el observador forma parte de aquello que es observado, la sociedad, y por ello, nunca puede ser totalmente neutral.

La observación social simple todos la realizamos, diariamente, con más o menos penetración, al entrar en contacto con fenómenos sociales.

Para que sea científica ha de reunir las siguientes condiciones:

1. Que sirva a un estudio ya formulado de investigación.
2. Que sea planificada y se realice sistemáticamente.
3. Que esté relacionada con proposiciones científicas más generales.
4. Que se empleen técnicas objetivas, es decir que aspire a observar, registrar e interpretar los hechos en forma tal que otras investigaciones puedan verificar sus hallazgos.

5. Que esté sujeta a algún control para la comprobación de su validez y fiabilidad.

Las formas de control que se suelen utilizar son: la realización de la observación por dos investigadores distintos a la vez y la discusión de los datos obtenidos con expertos y miembros del grupo. Asimismo, se recomienda tener en cuenta las reacciones, sentimientos y estado de ánimo de los observadores en el momento de la observación. (Ver ejercicios 70, 71)

SUS NOTAS

Según la definición expuesta, se pueden distinguir los siguientes rasgos o notas características de la observación simple:

1. Ser un procedimiento de recogida de datos que se basan en lo percibido por los propios sentidos del investigador. Por ello, se diferencia de la investigación documental y por cuestionario y entrevista, que se fundan en las percepciones contadas de palabra o por escrito de otras personas.

2. Consistir en el estudio de fenómenos existentes naturalmente o producidos espontáneamente y no provocados artificialmente como en los experimentos puros.

3. Ser un examen de fenómenos o acontecimientos actuales tal como son o tienen lugar en la realidad del momento presente, y no sobre hechos o acontecimientos pasados.

4. Que se realice con fines sociológicos, y con arreglo a los requisitos exigidos por la investigación sociológica científica, según su naturaleza descriptiva o explicativa. Esta condición distingue la observación sociológica de la realizada con otros fines, incluso científicos, no sociológicos. (Ver ejercicio 69)

5. En fin, la observación simple, como procedimiento de recogida de datos típico de las investigaciones sociales cualitativas, se caracteriza por ocupar en ella un puesto destacado la inducción, el paso de los datos a la teoría, de modo que ésta surge de aquélla, el enfoque holístico o global de las situaciones sociales y la comprensión o la pretensión de aprehender las intenciones en las acciones sociales.

TRATAMIENTO DE LOS DATOS DE LA OBSERVACION CUALITATIVA

Si bien esta obra se refiere más específicamente a los llamados métodos cuantitativos de la investigación social y aunque nada impide que

gran parte de su contenido pueda ser aplicado también, con las acomodaciones necesarias en las investigaciones sociales cualitativas, a continuación se enumeran, de acuerdo con S. J. Taylor y R. Bogdan (1986, 159 y ss.), los pasos básicos a seguir en el tratamiento de los datos de dichas investigaciones cualitativas.

ANALISIS

1. Leer detenida y repetidamente los datos contenidos en los cuadernos de campo, notas y en general documentos recogidos en la investigación.

2. Seguir la pista de temas, intuiciones e ideas surgidas durante dicha lectura.

3. Precisar los tópicos que aparezcan de manera recurrente en los documentos examinados.

4. Elaborar tipologías o esquemas de clasificación (V. pág. 415 de esta obra).

5. Formar conceptos, dando un nombre a las actividades, situaciones o fenómenos nuevos descubiertos.

6. Documentarse leyendo el material bibliográfico disponible que guarde relación con el asunto investigado.

CODIFICACION

1. Establecer las categorías pertinentes para efectuar la clasificación de los resultados obtenidos en el análisis.

2. Codificar o asignar a los mismos un código de identificación.

3. Clasificar cada uno de dichos resultados en las categorías en las que encajen mejor.

4. Discutir más detenidamente a qué categorías se pueden asignar los datos sobrantes del primer intento de clasificación.

5. Refinar todo el trabajo realizado, considerando conjuntamente y comparando entre sí los resultados obtenidos.

6. Relativizar las conclusiones inmediatas que se deriven de la investigación, teniendo en cuenta las circunstancias de su obtención y el contexto al que se refieren.

SUS CLASES

La observación presenta distintos grados y modalidades, aunque todos ellos son de gran interés para el sociólogo.

Existe, con relación a grupos sociales, además de la llamada no intrusiva que se realiza examinando por ejemplo, restos, basuras, graffiti, y en general, vestigios de las actuaciones sociales, una observación participante que se realiza enrolándose en el grupo observado y una observación distante que tiene lugar manteniéndose al margen de éste, aunque no tan alejado que se impida la observación pretendida. Además, según su grado de nivel científico, se pueden distinguir los tipos de observación siguientes: Precientífica, global, descriptiva y explicativa.

OBSERVACION PARTICIPANTE

Este tipo de observación tiene lugar, como se ha dicho, cuando el investigador se mezcla con el grupo observado y participa en sus actividades más o menos intensamente.

Si la participación es más prolongada y se realiza conviviendo con el grupo observado, durante un período más o menos largo de tiempo, entonces se tiene la observación antropológica o etnográfica, que

Si la participación es más prolongada y se realiza conviviendo con el grupo observado durante un período más o menos largo de tiempo, entonces se tiene la observación antropológica o etnográfica, que constituye una de las técnicas más importantes en estas ciencias sociales.

La observación primera, reportaje, tiene relación con el reportaje periodístico. Se diferencia de éste porque debe reunir las condiciones indicadas para que una observación sea científica. Además siempre se debe orientar a detectar aspectos de interés sociológico, previamente fijados.

La observación participante antropológica o etnográfica, se realiza conviviendo con las tribus o comunidades a investigar. Los estudios antropológicos existentes son, por lo general, resultado de la convivencia de los antropólogos con los aborígenes estudiados, a veces durante años. Primero se han de ganar su confianza y dejar que pase algún tiempo para que pierdan el miedo y obren naturalmente delante del investigador. Después deben ir observando y tomando notas de todas las particularidades de su vida en el orden social, político, cultural y económico.

No es posible desarrollar aquí las particularidades de este tipo de observación, pero si se puede recomendar la obra de S.T. Bruyn, *La perspectiva humana en Sociología*, publicada por la Ed. Amorrortu.

Esta obra considera a la observación participante, como ejemplo del tipo de investigación cualitativa, en contraposición a la cuantitativa, es decir, aquella que pretende penetrar y comprender el sentido íntimo de un fenómeno o realidad social, más allá de sus aspectos externos y cuantitativos.

En este sentido, se contrapone a la encuesta, que se suele considerar como forma típica de la investigación cuantitativa.

Este procedimiento también se puede utilizar para estudiar comunidades de pueblos avanzados culturalmente tales como aldeas, suburbios, etc., e incluso familias. Ejemplo respecto a familias son los estudios de O. Lewis, *La cultura de la pobreza* y *Los hijos de Sánchez*. Es famosa también la investigación de W.F. White, *Street corner society*, efectuado en un barrio popular de USA, formado por inmigrantes italianos y sus hijos. Igualmente la de Wilkie, *Village in the Vaucluse*, estudio sobre un pequeño pueblo cerca de Avignon. (Ver ejercicios 71, 76)

OBSERVACION PRECIENTIFICA

La observación precientífica, más bien que una técnica concreta, es una disposición de ánimo, una actitud ante la realidad social. Se puede decir que constituye el elemento fundamental del espíritu sociológico, es decir, de la aptitud para ver en todos los hechos de la vida los aspectos de interés social, de la capacidad para calificarlos y juzgarlos según categorías sociológicas y de la penetración para intuir los problemas de interés que presentan para su investigación científica.

Este tipo de observación debe constituir una cualidad permanente del sociólogo, básica no sólo porque toda investigación requiere un conocimiento lo más extenso y profundo posible de la realidad social, sino también como fuente de hallazgo de cuestiones científicas a investigar y de sugerencia de hipótesis significativas. Para adquirirlo se puede recomendar la lucha por conseguir una atención permanente a los hechos ofrecidos por la realidad, y el esfuerzo constante por aplicar a ellos y contrastarlos con categorías y principios sociológicos.

OBSERVACION GLOBAL

La observación global se puede considerar ya como una técnica sociológica, aunque sólo sea previa y preparatoria de una encuesta posterior. Aparece desarrollada y recomendada especialmente por J. Lebret, en su "Manual de Encuesta Social" (Madrid, Edit. Rialp). Se puede decir que consiste en un primer contacto general con el área, urbana o rural,

donde se piensa desarrollar una investigación para obtener un conocimiento básico de la misma en sus aspectos fundamentales, que sirva posteriormente de guía en la preparación y ejecución de la encuesta o estudio proyectado. Una forma de observación global es la preencuesta.

Su finalidad es percatarse lo más profundamente posible del paisaje y ambiente natural, económico, social y humano, a los que se va a referir el estudio, así como de las condiciones genéricas que caracterizan a todo universo a estudiar de cualquier tipo que sea.

A esta técnica se puede asimilar, el punto de vista en que insiste especialmente la sociología marxista de la necesidad de conocer previamente a toda encuesta los diversos intereses y conflictos de clase en la población investigada. Este es el medio para poder detectar las fuerzas sociales que actuan en ella y obtener una visión dialéctica de la misma.

Esta técnica, aunque no muy generalizada, se debe calificar de fundamental, pues un presupuesto básico de cualquier investigación es tener un conocimiento global del campo que va a abarcar. Se puede aplicar aquí lo que dice Caplow (1927, pág. 163). "Nada puede reemplazar un contacto directo del observador con su campo de estudio, ninguna técnica es capaz de sugerir tantas ideas nuevas". (Ver ejercicio 77).

OBSERVACION DESCRIPTIVA

Los otros tipos de observación, descriptiva y explicativa, constituyen técnicas de investigación sociológica independiente, y no meramente preparatorias como la observación global, aunque se utilizan también como accesorias o complementarias de otras técnicas. La distinción entre ambas radica en la finalidad descriptiva o explicativa de la investigación proyectada. También radica en el objeto de estudio, pues mientras la primera se refiere más bien a grupos o fenómenos sociales en conjunto (observación general de una fiesta de sociedad, de una manifestación publica, de una banda de jóvenes) en la explicativa es más reducido el fin de la investigación en cuanto tiende a comprobar determinadas hipótesis formuladas previamente.

La distinción entre ambos tipos de observación se refleja en el diverso rigor científico, exigida por una y otra.

En el caso de la observación descriptiva, como en los estudios de investigación de este tipo en general no es siempre necesario establecer una hipótesis explícita previa. Puede bastar determinar únicamente los aspectos del campo de observación que van a ser objeto de nuestra aten-

ción, o, en términos más concretos, especificar las variables sobre las que pretendemos obtener información.

La forma de proceder en este caso puede ser la siguiente:

1. Decidir el campo de observación y definirlo del modo más preciso posible.

2. Previo análisis de sus aspectos principales, es muy aconsejable a todos, pero especialmente a principiantes, elegir entre estos aspectos los que se crean de mayor interés limitando a ellos la investigación.

3. Especificar las dimensiones de los aspectos seleccionados y las variables empíricas o indicadores de estas dimensiones. Estas variables deben ser memorizadas por el observador.

Un ejemplo ya clásico de categorías e indicadores para la observación directa de la dinámica de grupos, son las establecidas por Bales después de pacientes experimentos sobre los tipos de comportamiento en las reuniones de grupo. Son estas:

1. Muestra solidaridad, anima. 2. Aparece relajado. 3. Aprueba. 4. Sugiere. 5. Da una opinión. 6. Da información. 7. Pide opiniones. 8. Pide una evaluación, una opinión. 9. Exige una orientación. 10. Rehusa, desaprueba, no participa. 11. Manifiesta una tensión. 12. Manifiesta agresividad.

4. Construir como instrumento de observación un cuadro de recogida de datos. En él, además de figurar los hechos descriptivos básicos de la unidad de observación: nombre, domicilio, situación, extensión, etc. se comprenden columnas en las que anotar la frecuencia e intensidad en su caso de los indicadores elegidos.

Este instrumento de observación debe ser ensayado o probado antes de empezar la observación, para comprobar si es adecuado, válido y seguro.

La anotación de las observaciones debe realizarse en el momento de la observación o si no es posible, después, pero lo antes que se pueda.

En la utilización práctica del instrumento de observación se plantea el problema de cómo lograr recoger la secuencia temporal de las interacciones. Hay diversos procedimientos. Acaso el más simple sea cambiar de hoja de observación cada cinco minutos, por ejemplo. (Ver ejercicios 73, 75)

OBSERVACION EXPLICATIVA

Según Mayntz y colaboradores (1975, p. 123), la observación explicativa, que el llama sistemática, se distingue de la no sistemática, por las siguientes características:

1) Sobre la base de una teoría explícita se aislan conceptualmente distintas clases de variables de comportamiento y de variables que determinan el mismo.

2) El proceso de observación se estructura en categorías de observación definidas con precisión, sobre las que se ordenan todas las observaciones relevantes. A este respecto, las variables han de ser operacionalizadas por medio de indicadores accesibles a la observación directa. Una dificultad especial se presenta frecuentemente al determinar las unidades de observación, que generalmente son comportamientos, debiendo fijarse al efecto lo que pudiera llamarse una acción "completa o cerrada"

3) Las condiciones situacionales de partida son sometidas a control a fin de posibilitar observaciones comparables, o por lo menos han de ser normalizadas o especificadas claramente

En el primer caso esto comporta por ejemplo la elección del grupo investigado, el señalamiento de las actividades que tienen que realizar y el establecimiento de las condiciones ambientales donde van a actuar, según criterios científicos.

Teniendo en cuenta estas características especiales y acomodándolos a ellas, los puntos acabados de indicar para la observación descriptiva pueden ser válidos para la explicativa. (Ver ejercicio 74).

EJERCICIOS

69

Dados los temas que se detallan a continuación, se pide señalar razonadamente su naturaleza y especialmente si se trata de verdaderas observaciones sociológicas directas simples e indicar en su caso su tipo:

A) El examen, según un plan previamente trazado y con objeto de ver la relación entre las técnicas de imprenta y las condiciones sociales de las épocas correspondientes, de la forma de impresión, presentación y, en general, condiciones externas de una serie de revistas y publicaciones pertenecientes a distintos períodos históricos.

RESPUESTA

Por su finalidad es una investigación sociológica, y aunque se refiere a documentos históricos, se puede considerar como una observación directa simple en cuanto pretende no el estudio de los hechos y datos contenidos en ellos, sino el análisis de sus condiciones objetivas externas.

B) La descripción detallada y pormenorizada de una manifestación política realizada por un gran semanario, tres días después de su celebración.

RESPUESTA

En este caso no existe observación sociológica, por su carácter y destino periodístico. Además, la descripción se realiza tres días después del suceso, con lo que se infringe una de las normas fundamentales de la observación sociológica, la de reseñar lo antes posible los hechos observados.

C) El estudio de los aspectos sociales de las relaciones económicas en la primera mitad del siglo XX, según los documentos y obras mercantiles de la época.

RESPUESTA

Tampoco hay aquí una observación sociológica directa simple en cuanto no versa sobre fenómenos actuales, sino históricos, y su fuente son los documentos, y no los acontecimientos. En realidad se trata de una observación de tipo documental.

D) El estudio sin ser visto de la interacción social de los niños de primer grado de una escuela, para comprobar la hipótesis de la existencia de liderazgo entre ellos.

RESPUESTA

Esta sí es una observación sociológica directa simple por estudiar un hecho espontáneo y cuando tiene lugar. Su tipo es explicativo en cuanto pretende comprobar una hipótesis y distante, porque el observador actúa al margen del grupo, sin participar en él.

E) El estudio de las reacciones de un grupo de niños seleccionados previamente a la proyección de una película policíaca escogida de antemano.

RESPUESTA

Esta investigación no es una observación sociológica directa simple, pues no es espontánea al ser seleccionado previamente el grupo y someterle a un estímulo dado, la proyección de una película. Se trata más bien de un tipo de observación experimental.

F) El examen previo de un suburbio en el que se piensa realizar una investigación sociológica.

RESPUESTA

En este caso existe una observación directa de carácter global, dirigida a conocer por experiencia propia un campo futuro de investigación.

G) El estudio de las formas sociales de vida de un grupo hippy mediante la convivencia con sus miembros durante tres meses.

RESPUESTA

También este estudio es una observación directa, pues estudia un fenómeno espontáneo, en su medio natural. Su tipo es participante, en cuanto se dice que el observador convive con el grupo.

H) La observación con propósitos descriptivos de sus aspectos sociales de las ceremonias de una boda típica rural, en la que el observador figura como invitado.

RESPUESTA

En una observación directa de tipo no estructurado, dado su fin descriptivo, y del propio grupo, en cuanto el observador participa como invitado en la ceremonia.

70

Analícense comparativamente las dos descripciones de una revuelta callejera y determínese razonadamente cuál de ellas es aceptable científicamente y cuál no.

PRIMERA

Hacía un tiempo muy bueno.

Su excitación mostraba que había participado en la revuelta.

Me tranquilizó con una sonrisa amistosa.

Iba vestido como un minero.

Los manifestantes estaban de pésimo humor; le había perdido el miedo a la policía.

SEGUNDA

Lucía el sol y la temperatura a la sombra era de 21°C.

La sangre caía sobre su cara y hablaba rápidamente, describiendo la revuelta.

Me miró y se sonrió.

Llevaba un mono azul, una chaqueta oscura, de cuero, y zapatos marrones del ejército.

Una decena de manifestantes hicieron frente a la policía, gritando sin cesar: "¡los grises, a sus casas!" y otras frases inaudibles.

RESPUESTA

De las dos descripciones no es aceptable la primera, porque en todos sus párrafos se encuentran valoraciones, juicios y deducciones subjetivas, al contrario de la segunda, en la que se detallan objetivamente y con pormenores concretos los hechos.

Por ello, en la primera descripción no se cumple la condición de que habla Caplow en su obra *La investigación sociológica*, de la cual están tomados los párrafos de las descripciones propuestas en el ejercicio, según la cual "las opiniones, las hipótesis inverificables o las observaciones sobre el carácter y la personalidad de los sujetos deben ser eliminados" (pág. 171).

71

En una observación sociológica directa realizada se han adoptado las siguientes medidas:

a) La observación y anotación de los hechos por dos observadores distintos.

b) En el informe de los hechos observados ambos observadores cuidaron de separar los hechos y las interpretaciones que éstos les habían sugerido.

c) Antes de su redacción definitiva dieron a conocer y discutieron las anotaciones tomadas con algún experto y también con miembros del grupo observado.

d) Ambos se preocuparon de exponer sus propias reacciones, sentimientos y estados de ánimo durante el tiempo de la observación.

Se pide indicar razonadamente si son correctas o no estas medidas y, en su caso, a qué finalidad pueden responder.

RESPUESTA

Todas estas medidas son admisibles, pues no se oponen a la naturaleza y condiciones de la investigación. La finalidad a que responden es lograr un cierto control de la validez y fiabilidad de las observaciones. En éstas, como se trata de casos individuales espontáneos, no reproducibles, en realidad no existe ningún procedimiento de verificación propia. Para paliar este inconveniente se recurre a estos expedientes y a otros similares a fin de lograr una cierta comprobación del trabajo efectuado.

72

Supuesto que se tiene que asistir a una fiesta patronal típica de una comunidad local, se ha proyectado realizar un estudio de observación, de tipo descriptivo, de los rasgos de interés social de este festejo popular.

Se pide: 1. Determinar las unidades de observación y la variable general a observar. 2. Señalar las dimensiones de esta variable. 3. Los indicadores a que se atendería principalmente, a fin de recoger información. 4. Reseñar los instrumentos que se podrían emplear para la recogida de datos y la forma de su disposición y utilización.

RESPUESTA

1. Las unidades de observación serían cada tipo de festejos: misa mayor, procesiones, bailes públicos, juegos y competiciones, fiestas de sociedad. La variable general a observar en todas ellas es sus aspectos de significación de social.

2. Como dimensiones principales de esta variable se pueden enumerar las siguientes: a) El elemento demográfico; b) el marco ecológico; c) factores culturales; d) estructura social; e) relaciones y comportamientos sociales; f) dinámica social del festejo.

3. Entre los indicadores de estas dimensiones sobre los que recoger información se encuentran:

a) Elemento demográfico: Composición de los que intervienen en el festejo por edades, sexos, profesional, según el estado civil, etc.

b) Marco ecológico: Condiciones geográficas y climáticas generales de la localidad. Lugar del festejo: aire libre, sitio cerrado. Sus caracteres ambientales: temperatura, ventilación, iluminación, etc.

c) Factores culturales: Vestidos usados. Atuendo y aderezo de los participantes. Instrumentos utilizados. Músicas, cantos, bailes. Normas y costumbres que regulan el festejo. Valores y creencias a que responden.

d) Factores estructurales: Existencia de grupos formales e informales. Papel de los directores o líderes. Distinción de clases. Diferentes funciones asignadas a los partícipes.

e) Comportamiento social: Acciones sociales más significativas de los actores y espectadores. Impacto y respuesta que provocan. Relaciones entre los organizadores o líderes, en su caso, y el público. Relaciones entre los actores y espectadores, y de éstos entre sí. Manifestaciones de alegría y enfado: aplausos, protestas, etc.

f) Dinámica del festejo. Cambios de ritmo a lo largo de su duración. Modificaciones en la actitud de los participantes. Variaciones sentimentales de actores y público. Efectos producidos por el festejo después de su terminación. Satisfacción o decepción de la gente por él. Comentarios de la gente. Tensiones, disputas y conflictos producidos a lo largo del festejo.

4. Instrumentos de recogida de datos. Podrían ser dos: Primero, un block manejable para tomar nota durante las ceremonias, y segundo, un conjunto de fichas o un cuaderno en el que estuvieran dedicadas una o más páginas a cada categoría. Durante los festejos se irían tomando apuntes en el block de notas, cuidando de rotularlo según las categorías previstas. Después de terminar, o en los ratos de descanso, se reseñarán en el cuaderno alfabético o en las fichas, teniendo en cuenta las notas recogidas, los puntos interesantes sobre cada categoría.

73

Supuesto que se quiere estudiar empleando el método de observación directa las reacciones a los actos de agresión dentro de un grupo de diez niños que juegan vigilados por un supervisor, se pide:

1. Determinar las unidades de observación y las variables.

2. Definir operativamente la unidad de observación.

3. Especificar las dimensiones e indicadores de las variables en cuestión.

4. Construir un cuadro para registrar en el campo de observación la procedencia de los actos de agresión y las reacciones a ellos.

RESPUESTA

1. Hay una sola variable en este ejercicio, la reacción a los actos de agresión. Las unidades de observación son estos actos.

2. La unidad de observación: acto de agresión se podría definir como "todo acto violento verbal o físico dirigido a otro niño que tiende a causarle un daño corporal o moral".

Como se ve, en esta noción se pueden distinguir dos tipos de agresión principales: los físicos y verbales. Por simplificación se podría limitar la observación a los físicos, si bien se podría también dedicar un cuadro a las agresiones de tipo físico y otro a las de tipo verbal y después establecer comparaciones entre ellos.

3. En la variable reacción a la agresión se podrían distinguir dos dimensiones: física y sentimental. Como indicadores de reacción física están: las lágrimas y la contraagresión física y la verbal. Como indicadores de reacción sentimental se señalarán la ira, la sorpresa y la indiferencia.

Por razón de sujeto, la reacción a la agresión se puede distinguir según se miren sus efectos en el sujeto pasivo, en los otros compañeros o en el vigilante.

4. El cuadro se formaría dedicando las filas una para consignar el número de los niños, sujetos pasivos de la agresión, y otras para cada uno de los tipos de reacción a la agresión separadamente y para la ayuda en su caso de los otros niños al agredido y también separadamente del vigilante. En cuanto a las columnas, se dedicaría a cada niño un espacio, con indicación en él de su número y sexo, de cuatro o cinco columnas, para la anotación de otros tantos actos de agresión iniciados por cada niño. (Vid. sobre una observación de este tipo, Wiseman, 1971, pág. 21.)

74

Para comprobar en la realidad la hipótesis de que los "nuevos miembros de un grupo pequeño adoptan una posición de inferioridad respecto de los veteranos" se ha proyectado realizar una investigación sociológica en una pequeña clase de alumnos de primer grado de primaria, en la que se incorporaron cinco miembros nuevos. Se pide:

1. Distinguir en la hipótesis las unidades de observación y las variables.

2. Operacionalizar las variables, determinando sus dimensiones e indicadores.

3. Indicar la forma de construir el cuadro para anotar los resultados de la observación, según los indicadores elegidos y el número de niños objeto de la observación.

RESPUESTA

1. Unidad de observación: cada uno de los nuevos miembros del grupo. Variable: conducirse como inferiores respecto a los alumnos antiguos.

2. En esta variable se pueden distinguir dos dimensiones: conducta de inferioridad pasiva y activa. Como indicadores de la primera dimensión, conducta pasiva,

están: aguantar agresiones, aguantar insultos, aguantar burlas, no atreverse a juntar con los veteranos, no atreverse a hablar estando en grupo con ellos, etc. Entre los indicadores de la segunda dimensión, conducta de inferioridad activa, se pueden enumerar: hacerles obsequios, hacerles pequeños servicios, sonreirles amablemente, procurar juntarse con ellos, escucharles con atención, mostrar admiración por sus actos y palabras, etc.

3. El cuadro se podría construir de modo similar al del ejercicio anterior, dedicando las filas a los indicadores elegidos y las columnas un espacio con cuatro o cinco columnas a cada uno de los niños objeto de la observación.

75

Supuesto que se tiene que realizar una investigación sobre las condiciones de seguridad en el trabajo de las industrias de muebles existentes en una localidad centro productor de ellos, se pide:

1. Determinar la unidad de observación y la variable a investigar.

2. Especificar las dimensiones de esta variable y sus indicadores.

3. Indicar cómo se construiría el cuadro de observación.

RESPUESTA

1. Unidad de observación: la fábrica o el taller, en el caso de que una fábrica tenga varios. Variable a investigar: las condiciones de seguridad.

2. En dicha variable, según los elementos que forman la fábrica o el taller, como unidad de producción, se pueden considerar las dimensiones siguientes: los trabajadores, el local, las máquinas y materias primeras, las instalaciones accesorias.

3. Indicadores o variables empíricas:

a) Trabajadores. Uso de pertrechos y atuendos de protección: casco, guantes, botas especiales, gafas, protecciones especiales en el vestido, etc.

b) Local. Materiales con que están construidas las paredes, el piso y el techo. Número de salidas. Tipo: directas o no a la calle. Salidas de socorro. Número y tipo de ventanas. Su clase y protección en ellas, etc.

c) Maquinaria y materias primas. Máquinas y motores: tipos de protección. Sistema de entretenimiento. Frecuencia de revisión de su funcionamiento. Especialización del personal que las maneja. Materias primas: orden de almacenamiento, forma de custodia y conservación de las materias inflamables y explosivas, medidas de seguridad en ellas, etc.

d) Instalaciones accesorias:

Iluminación: tipo. Instalación eléctrica: clase de cables, existencia de cortocircuitos. Extintores de incendios: tipos, cantidad, colocación. Calefacción: sistema, protección. Ventilación: tipo, condiciones. Existencia de botiquín, etc.

3. El cuadro constaría en primer lugar de un encabezamiento para consignar en él los datos básicos sobre la empresa y el taller: Nombre, domicilio, número de trabajadores, tipo de local, su extensión, etc. En el cuerpo del cuadro las filas se dedicarían a los diversos indicadores elegidos, destinándose una o varias columnas a la anotación de las observaciones realizadas.

76

Un estudiante de Sociología se ha inscrito en un campo de trabajo de minería durante las vacaciones de verano con objeto de estudiar, mediante la observación participante, las normas por las que se rige el grupo de trabajo en que se le encuadre.

Supuesto lo anterior, se pide:

1. Determinar la unidad de observación y la variable a estudiar.

2. Especificar las dimensiones de esta variable y los indicadores a que se debería prestar atención especial.

3. Indicar los instrumentos de observación y las formas de su utilización.

RESPUESTA

1. Unidad de observación: el grupo de trabajo. Variable: las normas vigentes en el grupo.

2. Como dimensiones de esta variable se pueden distinguir:

a) Las normas referentes a la ejecución del trabajo: b) Normas sobre la relación con otros estamentos de la empresa y la dirección; c) Normas sobre relaciones internas del grupo; d) Normas referentes a las relaciones con personas extrañas al grupo y a la empresa; e) Normas punitivas relativas a las sanciones por infracción de las reglas del grupo.

Entre los aspectos de cada una de estas dimensiones se encuentran los siguientes: a) Normas sobre el ritmo, rendimiento, forma de ejecución del trabajo, cuidado y trato de los instrumentos y materiales. b) Normas referentes a las relaciones con otros estamentos y la dirección: formas de conducta alabadas y criticadas con la dirección, con los ingenieros, con los capataces, con los empleados. c) Normas sobre recepción de nuevos miembros, sobre socialización y adaptación a las pautas del grupo, sobre ayuda mutua, sobre solidaridad en cuestiones de la empresa. d) Normas sobre relaciones con trabajadores de otros grupos de la misma empresa, con trabajadores de otras empresas, con los familiares de los miembros del grupo, con personas de otro sexo, etc. e) Normas referentes a sanciones por exceso o defecto de rendimiento, de ritmo y en la forma de ejecución del trabajo y trato de instrumentos y materiales; sanciones por alcahuetismo, sanciones por falta de solidaridad y compañerismo, etc.

3. Block de notas para apuntar datos de interés durante los ratos de descanso y diario para reseñar en él cada día las observaciones, experiencias e informaciones recogidas sobre el objeto de la investigación, cuidando de transcribirlas de manera ordenada y clasificada según los diversos aspectos indicados a observar. Este diario servirá de base para la elaboración del informe final, después de concluida la investigación, en el que se expondrán los resultados de ésta.

77

Se ha proyectado realizar el estudio de las condiciones socioeconómicas de un suburbio, y supuesto que, con vistas a esta investigación, se nos encarga realizar la observación global previa de dicho suburbio, se pide indicar los aspectos generales

a que se debería dedicar especial atención en la observación y la manera práctica de proceder en ella.

RESPUESTA

Siguiendo el *Manual de Encuesta Social* de Lebret cit. (págs. 23 del vol. 2 y 38-39 del vol. 1), los puntos significativos sobre los que se debería centrar la atención y el modo de actuar, son los siguientes:

1. Asimilación del paisaje natural, obteniendo una visión de conjunto del mismo desde una colina, torre o punto alto que lo domine. Si es posible, se procurará sobrevolar el territorio en que se asienta para descubrir la estructura esencial de la agricultura y la industria, la relación del suburbio con la ciudad y el campo y su posición respecto a ellos. En esta primera actividad no debe olvidarse el mapa de la región para contrastarlo con la realidad y procurar su mejor comprensión.

2. Hay que recorrer las calles comerciales y los mercados, observando a la gente que entra y sale de ellos, la densidad y tipo de vehículos que circulan, las aglomeraciones de tráfico y de peatones, el aspecto de éstos y su indumentaria.

3. Es preciso visitar también las zonas periféricas para captar las características y amplitud del paisaje de fábricas, dedicando especial atención a los sectores más míseros, visitando algunas chabolas y las zonas acomodadas, si las hay.

4. Deberá estudiarse la naturaleza y condiciones externas de las fábricas y su emplazamiento. Es necesario visitar una fábrica por lo menos de cada industria principal. También se tomará nota de la calidad y característica de los edificios.

5. Deberá recorrerse en coche o a pie las vías de comunicación del suburbio con el centro de la ciudad, a fin de darse cuenta de la densidad y tipo de circulación, de los embotellamientos y del tiempo necesario para trasladarse de un lugar a otro. Esta observación tiene particular interés a la hora de la salida de las fábricas y de mayor tráfico de los camiones de carga.

6. Es necesario observar también a este respecto los enlaces (transportes colectivos) con la ciudad, sus condiciones y regularidad de funcionamiento.

7. Igualmente hay que procurar descubrir genéricamente el grado de dependencia y autonomía del suburbio respecto a la ciudad en los aspectos industrial, comercial, hortícola, residencial, de centros de educación.

8. El panorama humano se descubre observando trabajar a la gente en los comercios, mercados, talleres, a los caminantes que pasan por las calles, a los obreros a la salida de las fábricas, a los que entran y salen de los grandes almacenes, asistiendo a los oficios religiosos o confundiéndose en las manifestaciones populares y entre las multitudes que asistan a ferias y mercados.

9. Además, es preciso establecer contactos con las diversas capas de población en cafés, bares, restaurantes, lugares de entretenimiento y círculos familiares. A este respecto no hay nada comparable a las comidas en el seno de algunas familias representativas de las diversas capas sociales. Fijarse en sus preocupaciones, su nivel cultural, sus formas de acogida y de cortesía. Igualmente es muy importante visitar algunas personas significativas e interrogarlas sobre cuestiones de interés respecto a las actividades agrícolas, artesanas, industriales y comerciales.

15. OBSERVACION EXPERIMENTAL

NOCION

Como ya se ha indicado en la introducción a la observación simple, el experimento es en sí una observación directa, en cuanto se basa en el estudio, mediante el empleo de los sentidos del investigador, con o sin ayuda de aparatos técnicos, de hechos y fenómenos de interés social, cuando tienen lugar en la realidad.

Sin embargo, no es una observación simple, en sentido estricto. Primero porque consiste en el estudio de grupos y fenómenos, y ésta es sin duda su nota distintiva, preparados o manipulados, es decir, controlados de alguna forma por el investigador. Segundo porque el experimento es el instrumento de investigación causal por excelencia, a diferencia de la observación que, generalmente, es simplemente descriptiva. Y tercero porque implica la contraposición bien de dos grupos o fenómenos similares, salvo en un influjo producido o en una situación que se da por sí misma, bien del mismo grupo o fenómeno esta vez en dos tiempos distintos después de haber sufrido un impacto determinado. Precisamente la técnica del experimento se basa en la producción o preparación de esta contraposición o contraste, de tal modo que en ella consiste principalmente la manipulación del objeto de la investigación que supone el experimento.

En su desarrollo pleno la técnica del experimento comprende las siguientes fases: 1) Preparación de dos grupos o situaciones idénticas o lo más similares posibles. 2) Observación o medición inicial simple de ambos grupos. 3) Modificación de uno de los grupos o situaciones, haciendo incidir sobre él una variable (estímulo). 4) Nueva observación y medición de ambos grupos, especialmente en cuanto a la variación producida por el estímulo en el grupo o fenómeno sometido a él. 5) Compara-

ción de los resultados obtenidos en los grupos mediante la medida de las diferencias entre la situación final e inicial de cada uno de ellos.

Los elementos básicos que intervienen en la observación experimental son a) el ambiente en general en el que se efectúa el experimento; b) los grupos que se contraponen o comparan y c) el estímulo o variable experimental que se hace incidir sobre el grupo experimental.

Cuando todos estos elementos están bajo el control y manipulación del investigador, se tiene el experimento que se puede llamar puro o totalmente controlado.

El control del ambiente se realiza mediante el aislamiento experimental, es decir, efectuando el experimento en un laboratorio donde el investigador puede controlar o mantener constantes las variables ambientales: luz, temperatura, presión atmosférica, etc., que pueden influir en los grupos sujetos a experimento. Este control del ambiente es, sin duda, importante en los experimentos físicos y biológicos, pero en los experimentos sociales presenta el gran inconveniente de constituir el laboratorio un ambiente artificial, que como tal puede influir física y psíquicamente de manera importante en las personas sujetas a experimento.

Cuando no se pueden controlar algunas de las variables ambientales se recurre a su aleatorización. Esta técnica tiene aplicación en el caso en el que se puedan realizar las repeticiones del experimento que sean necesarias. En este caso, se hace que las variables incontroladas intervengan en cada experimento con valores distintos, altos, medios, bajos, lo que permite detectar su influencia en el experimento y eliminar la posibilidad de influjo sistemático encubierto.

El control de los grupos con los que se realiza el experimento tiene lugar, principalmente, por medio de la igualación de los grupos experimental y de control, de la que se trata después, respecto sobre todo a las características básicas en orden a los objetivos del experimento.

Por último, el control del estímulo o variable independiente es la facultad del investigador de hacer incidir tal estímulo sobre el grupo que parezca conveniente y en el grado y momento que le parezca oportuno.

Este control del estímulo es, sin duda, como se ha indicado, lo más típico del experimento. Esto explica que D. Campbell y J.S. Stanley (1973, p. 9) entiendan por experimento "aquella parte de la investigación en la cual se manipulan ciertas variables y se observan sus efectos sobre otras".

En el experimento las variables independientes son las que el investigador hace variar y las dependientes son las que observa para ver si varían a causa de la variación provocada por el investigador en las variables independientes.

También principalmente en los experimentos, pueden intervenir otras variables que Galtung llama relevantes e irrelevantes. Las primeras son importantes en relación al fenómeno estudiado y, por ello, el investigador debe controlarlas manteniendolas a lo largo del experimento sin variación, a fin de que no puedan influir en la variación a observar de las variables dependientes.

Las irrelevantes son las que se puede hacer caso omiso de ellas, bien porque se estima que su influencia es despreciable, o bien porque esta influencia es aleatoria y se anula en conjunto.

El experimento es la técnica de investigación científica más poderosa. Como dicen Yule y Kendall, en su obra *Introducción a la estadística matemática,* deriva su poder de la habilidad del experimentador para reemplazar los complejos sistemas de causalidad reales por sistemas sencillos. Su fin es posibilitar el estudio del fenómeno en las condiciones elegidas por el investigador.

Por ello, los experimentos bien hechos sirven para descubrir si una variable es independiente respecto de otra, en el sentido indicado de que influye o es causa de la variación observada en la variable dependiente.

En las técnicas no experimentales, llamadas correlacionales, ésto solo se puede establecer con una cierta probabilidad, mediante el estudio matemático estadístico de las correlaciones entre las variables y utilizando técnicas lógicas.

Sobre todo se basan en el experimento las ciencias físico-químico-naturales. En sociología es también la técnica ideal, pero sólo se puede emplear por lo general de modo imperfecto. Motivos humanos y morales impiden en muchos casos su aplicación a los grupos sociales y, aun supuesta su admisibilidad moral, no es posible por lo general en las ciencias sociales el control científico riguroso del ambiente, de los grupos y de la variable experimental como exige el experimento.

No obstante, conviene señalar con E.B. Wilson (1953, p. 56) que "es raro que solo un experimento sea considerado como suficiente; usualmente las repeticiones se estiman necesarias en orden a contrastar los resultados y también como base de estimación de la precisión de los resultados obtenidos. Este proceso de replicación es especialmente necesario

cuando la clase bajo estudio no está definida demasiado precisamente y está por eso sujeta a amplias variaciones individuales". (Ver ejercicio 94).

TIPOS DE EXPERIMENTO

Los experimentos pueden clasificarse según el ambiente en que se realizan, según su desarrollo o perfección y según su naturaleza.

Según el ambiente, se distinguen los experimentos de laboratorio y experimentos sobre el terreno o de campo.

Como escribe Caplow (1972, p. 187), un laboratorio es un lugar cerrado, sometido al control de un investigador y dotado de un equipo que le permite efectuar las manipulaciones y mediciones necesarias. El ambiente de laboratorio es, por definición, artificial y como tal sólo reproduce aquellos aspectos del ambiente natural que el investigador quiere resaltar. Además nunca se puede estar seguro que este proceso de selección no introduzca modificaciones del contorno incontroladas.

La casi totalidad de las investigaciones sobre procesos sociales efectuados en laboratorios son investigaciones de psicología social que afectan tanto a la sociología como a la psicología.

El laboratorio típico de psicología social consiste en una sala de dimensiones medias, equipada con mesas, sillas y pantallas, y con magnetófonos, cámaras y otros aparatos de registro. Además debe tener una sala de observación insonorizada separada de la sala de experimentación por un espejo que permita ver sin ser visto.

Los experimentos de campo o sobre el terreno son los que no tienen lugar en un laboratorio. Los experimentos de este tipo si fueran rigurosos no deberían diferir de los experimentos de laboratorio. También en ellos es necesario el control perfecto de la situación social de origen. Pero, como esto es muy difícil de conseguir, la realidad es que los experimentos sociales sobre el terreno son generalmente experimentos imperfectos. Un ejemplo de experimentos de esta clase es el realizado con personas alojadas en barracas. Se les dió solo a algunos de ellos viviendas nuevas para ver qué diferencias sociales se producían.

De acuerdo con el segundo criterio se pueden distinguir los cuasi-experimentos y los experimentos propiamente dichos.

Los cuasi-experimentos se consideran formados por aquel tipo de investigaciones en el que falta algún elemento importante, por ejemplo el grupo de control, o suponen un control incompleto sobre los elementos

señalados importantes en el experimento social: los grupos observados y el estímulo.

Por el contrario, experimentos propiamente dichos son aquéllos en los que no falta ningún elemento básico y el control del investigador alcanza tanto a los grupos experimentados como al estímulo o variable independiente.

Todos los tipos de experimentos expuestos se pueden clasificar de reales en cuanto implican la producción real, bien de manera espontánea o artificial, de un fenómeno social.

Sin embargo, han adquirido gran importancia últimamente los llamados experimentos simulados, que cada día tienen más aplicación en todas las ciencias , debido a la dificultad y coste de los experimentos reales.

Este tipo de experimentos, en oposición a los reales, no se basan en la observación de la incidencia de un influjo real determinado sobre una situación o grupo efectivos. Al contrario, se basan en el estudio de las consecuencias sobre una situación simulada, ficticia, a veces púramente teórica, de la modificación de algunos de sus elementos. La situación teórica viene a ser la situación o grupo experimental de partida y la modificación de sus elementos o factores se puede decir que constituye el estímulo del experimento simulado.

En este tipo de experimentos se basan en síntesis en lo siguiente. Se construye un modelo que como tal pretende ser una representación de la realidad. Entonces, en lugar de estudiar la realidad directamente se modifican algunas de las variables que forman el modelo y se estudian las transformaciones que se producen en el conjunto del modelo por tal variación. Los resultados obtenidos se pueden verificar simuladamente mediante series de valores que se obtienen por medio de los ordenadores.

Estos experimentos tienen aplicación sobre todo en estudios prospectivos que intentan prever el futuro. Son muy útiles en política social y económica, para prever las modificaciones que se producirán en la sociedad y en la economía si se actúa sobre determinadas variables. (Ver ejercicio 94).

Por último quedan los llamados experimentos ex-post-facto.

Se producen cuando falta la situación inicial y el control del estímulo, ya que actúa el investigador después de haber actuado la causa o variable independiente y de haberse producido los efectos consiguientes.

Se puede decir que de experimentos tienen muy poco, salvo un cierto control imperfecto que es posible en ellos de la elección de los miem-

bros del grupo que ha sufrido el impacto de la variable independiente y de los del grupo de control. Como señala el Profesor González Seara (1971) estos experimentos se deben utilizar con mucha cautela, pues las explicaciones ex-post-facto pueden utilizar los datos como mejor convenga. (Ver ejercicio 78)

OPERACIONES PREPARATORIAS DEL EXPERIMENTO

En cuanto a las operaciones preparatorias del experimento se pueden destacar: a) Análisis de la hipótesis y determinación del estímulo. b) Formación de los grupos experimental y de control. c) Medición de estos grupos, antes y después del estímulo. Y d) Construcción del modelo del experimento y significación de los resultados.

a) El análisis de la hipótesis no ofrece particularidad especial en relación a lo ya expuesto al tratar de ellas. El estímulo es precisamente la variable independiente, que bien se hace incidir en el grupo, o bien ha actuado ya en él naturalmente.

b) Para la formación de los grupos experimental y de control existen tres procedimientos principales. Los llamados control de precisión, control por distribución de frecuencias y control por azar. En el primero se realiza la elección de los miembros de ambos grupos, individuo por individuo. Seleccionado, por ejemplo, un miembro para el grupo experimental se busca otro que reúna las mismas circunstancias tomadas en consideración en el primero para el grupo de control, y así sucesivamente hasta completar el número previsto. En cambio, en el control por distribución de frecuencias esta igualación no se hace de modo invididual, sino globalmente, de modo que la composición de ambos grupos sea igual respecto a las características en cuestión. Por ejemplo, si el grupo experimental está formado por un 30 por 100 de individuos de hasta quince años de edad, un 50 por 100 de dieciséis a dieciocho y un 20 por 100 de dieciocho a veinte, el grupo de control deberá tener la misma composición por edades e igualmente en relación a las demás características del grupo.

El procedimiento de azar se puede utilizar para igualar aleatoriamente los dos grupos. Para ello se eligen dos muestras de igual número de sujetos de la misma población, según se ha indicado ya. También se puede emplear para realizar la asignación de los individuos a los grupos de control o experimental. En este caso, este procedimiento se puede combinar con cualquiera de los dos primeros.

Se ha de señalar, que la igualación ha de hacerse sobre todo respecto a variables independientes altamente relacionadas con la variable dependiente investigada.

c) Para la medición de los grupos después del experimento, y antes, en su caso, se aplicarán las técnicas existentes en Sociología, por ejemplo, escalas de actitudes, tests, entrevista, cuestionario, etc. Aquí es aplicable lo expuesto al tratar de la elección de los métodos a emplear.

d) El modelo de experimento, en el caso de ciertos diseños experimentales y cuasiexperimentales, es un diagrama formado por un cuadrado dividido en cuatro partes iguales que corresponden a las medidas antes y después del experimento de los grupos experimental y de control.

Como es obvio, cuando en el experimento solo hay un grupo, dos cuadros estarán vacíos, asi como cuando no hay medidas antes y sí sólo después. Esto se indica por una línea de trazos discontinuos.

La diferencia de las diferencias entre las medidas después y antes o entre las del grupo experimental y el de control, si no hay momentos antes, es el resultado del experimento, el posible efecto del estímulo o variable experimental X.

Su significación es simple. Si esta diferencia es elevada quiere decir que el experimento no revoca la hipótesis y lo contrario si es baja. Para ver si estas diferencias observadas son significativas estadísticamente, se puede emplear la prueba t o el análisis de varianza. (Ver ejercicios 79 a 82).

Otras cuestiones importantes respecto a la observación experimental son el diseño y la validez. Son tratadas en el capítulo 7.

EJERCICIOS

78

Dados los casos de investigaciones que se describen a continuación, se pide indicar razonadamente si son o no experimentos sociológicos y en caso afirmativo determinar a qué tipo de experimento corresponden.

1. El estudio del comportamiento de los niños de una escuela de párvulos mediante la observación sistemática de sus reacciones a los múltiples estímulos a que ven sometidos en la escuela, por varios investigadores a la vez, con el fin de lograr el control de las variaciones en la observación debidas al observador.

RESPUESTA

Esta investigación no es experimento porque no existe en ella el grupo de control, ni pretest o medición previa de los grupos investigados, así como tampoco control de los estímulos.

2. Los trabajos realizados por Sorokin para estudiar si, en igualdad de condiciones, el rendimiento del trabajo varía según el sistema de remuneración empleado: igual o desigual, individual o colectivo. Para ver la solución a este problema se utilizó niños de edad preescolar de tres a cuatro años, a los que se hizo realizar trabajos determinados apropiados a su edad. Como remuneración se empleó diversas clases de juguetes. En la remuneración colectiva, los juguetes no podían ser llevados a sus casas por los niños como si fueran suyos, sino quedaban en la sala de juguetes para uso y disfrute de todos los niños. En el caso de la remuneración individual el niño podía hacer lo que quería con el juguete. En la remuneración igual, los niños recibían juguetes idénticos. A cada cambio de tipo de remuneración se midió el trabajo ejecutado por el mismo grupo de niños por unidad de tiempo.

RESPUESTA

En este caso se trata de un cuasiexperimento del tipo de Campbell y Stanley (1973, p. 86) llaman de muestras cronológicas equivalentes.

En él se da la contraposición del mismo grupo consigo en cuanto es sometido a estímulos diferentes referentes al sistema de remuneración y objeto de diversas observaciones.

Su diseño sería del tipo siguiente: $X_1 \ O_1 \ X_0 \ O_2 \ X_1 \ O_3 \ X_0 \ O_4$.

3. Una investigación sociológica realizada mediante la aplicación de un cuestionario para conocer la relación de la clase social y del tipo de estudio con el grado de conocimiento sindical y la actitud hacia los sindicatos.

RESPUESTA

No existe en esta investigación ni experimento ni cuasiexperimento. En primer lugar, su finalidad es buscar si hay asociación entre determinadas variables y no una relación de causalidad entre ellas. En segundo lugar, no presenta grupo de control, pretest, ni siquiera estímulo o variable independiente.

4. El conocido estudio realizado para descubrir la eficacia en los hábitos sanitarios de la aplicación de un plan de higiene rural en una aldea de Siria en contraste con otras dos aldeas también sirias, no sometidas al plan, ni comunicadas con la anterior, pero de condiciones geográficas, demográficas, históricas, económicas, religiosas, domésticas, educativas, recreativas y sanitarias similares.

RESPUESTA

Se trata de un experimento propiamente dicho del tipo "diseño de grupo de control pretest-postest".

En él, por tanto, aunque no se dice en el enunciado, se debieron realizar cuatro observaciones: dos a la aldea experimental y dos a la aldea de control, una antes y otras después de someter a la aldea experimental al plan de higiene rural.

5. Los estudios de Carr y Angel en la Universidad de Michigan, que dividieron sus estudiantes en grupos de cuatro o más y asignaron a cada grupo un problema a resolver, tal como planear una velada social, organizar una excursión campestre de la clase, conseguir un acuerdo sobre algún punto, controlando las condiciones ambientales y observando sistemáticamente el desarrollo de la integración social en cada reunión dentro del grupo.

RESPUESTA

Según el enunciado, este estudio no es experimento ni cuasiexperimento porque falta el grupo de control, no se dice nada sobre la igualación de los grupos estudiados y al parecer sólo existe una medición de los mismos.

6. El estudio de Mandel sobre la relación existente entre la duración de la pertenencia a los boy scouts y la participación posterior de éstos en las actividades de su comunidad, realizado mediante la comparación de dos grupos similares en todo lo demás, excepto en que un grupo había pertenecido a los boy scouts un promedio de 1,3 años, mientras que el otro había pertenecido a ellos cuatro años.

RESPUESTA

No es un experimento propiamente dicho, ni tampoco un cuasiexperimento. Se trata de un experimento impropio, de los llamados ex-post-facto, que versa sobre situaciones reales pasadas que se observan después de haber actuado las causas y haberse producido los efectos.

En él se da únicamente la formación de dos grupos similares a efectos de la investigación, pero no existe pretest ni control alguno sobre el estímulo que ha actuado ya por sí mismo antes.

7. La investigación realizada por Lazarsfeld y Gaudet con dos grupos similares en número, edad, sexo, educación, nacionalidad, religión, etc., que se diferenciaban en que uno de los grupos estaba formado por personas colocadas y el otro por parados que buscaban empleo. A estos grupos se sometió a cuestionarios, entrevistas y escalas en orden a descubrir diferencias sociales y de personalidad entre ellos.

RESPUESTA

Esta investigación es como la anterior un experimento ex-post-facto. Por ello es aplicable aquí fundamentalmente lo acabado de exponer en la respuesta al caso precedente.

En este caso aparece claramente la imperfección de los experimentos sociales, sobre todo de este tipo, pues es evidente que en la situación de empleo o paro de una persona pueden influir muchos factores y que igualmente se podría tomar, en cierto modo, el paro como causa y las diferencias de personalidad como efectos en cuanto sin duda dicha situación influye también en estas diferencias de personalidad.

79

Partiendo de la hipótesis de que la información sobre los efectos perniciosos de las drogas influye en la aceptación por los jóvenes de su consumo, diséñese un experimento para contrastar esta hipótesis empíricamente, y para ello:

1. Analizar la hipótesis indicada y señalar en ella la variable dependiente y la independiente o estímulo.

2. Señalar qué procedimiento se podría utilizar para formar los dos grupos del experimento, el experimental y el de control, en un colegio de BUP y COU, para jóvenes de ambos sexos de la clase alta.

3. Indicar algún procedimiento concreto que se podría utilizar para hacer incidir el estímulo o variable independiente en el grupo de control.

4. Reseñar el tipo de experimento, formar su modelo y elegir las técnicas que se podría utilizar para observar o medir las actitudes y opiniones de ambos grupos antes y después del experimento.

5. Supuesto que entre los resultados obtenidos se encuentran los siguientes: Antes del experimento un 49 % del grupo de control y un 50 % del grupo experimental han confesado que les gustaría fumar marihuana. Después del experimento dicho porcentaje fue este: un 48 % en el grupo de control y un 25 % en el experimental. Señalar si se puede o no rechazar la hipótesis formulada.

RESPUESTA

1. Análisis de la hipótesis.

Unidad de observación: cada jóven.

Variable dependiente: aceptación por los jóvenes del consumo de la droga.

Variable independiente: información sobre los efectos perniciosos de la droga.

2. Se podría utilizar el procedimiento aleatorio, eligiendo al azar dos muestras estratificadas por cursos y sexo de la población estudiantil del colegio.

3. El estímulo, la información sobre la droga, se puede hacer incidir en el grupo experimental, sometiendo a éste, por ejemplo, a un cursillo de conferencias sobre el tema a cargo de especialistas en la materia: médicos, psicólogos, sociólogos, etc.

4. Se trata de un experimento propiamente dicho, del tipo, según el enunciado, "diseño de grupo de control pretest-postest.

El modelo del experimento, pues, es del tipo siguiente:

	Aceptación de la droga inicial	*Aceptación de la droga inicial*
Grupo experimental	50 %	25 %
Grupo de control	49 %	48 %

La técnica a emplear para hacer las mediciones u observaciones, antes y después del experimento, es, de acuerdo con el enunciado, el cuestionario.

5. Resultados:

Variación en la aceptación de la droga en el grupo experimental: 25 % - 50 % = = -23

Variación en la aceptación de la droga en el grupo de control: 48 % - 49 % = - 1 .

Diferencia en la variación de ambos grupos: 23 % - 1 % = 22

Es una diferencia elevada, por tanto se puede seguir manteniendo la hipótesis de partida según la cual la información influye en la aceptación de la droga por los jóvenes estudiantes.

Para ver si esta diferencia es significativa en relación a toda la población del Colegio de la que proceden las muestras que formaban los grupos experimentales, se puede utilizar el test estadístico t.

80

Partiendo de la hipótesis de que los estudios de sociología influyen en las actitudes sociales de los estudiantes que empiezan esta carrera, y supuesto que se piensa proyectar un experimento para comprobar la validez y el sentido de esta hipótesis, se pide: 1. Analizar la hipótesis distinguiendo la unidad de observación, las variables y el tipo de experimento, indicando la variable que actúa de estímulo. 2. Señalar cuáles y cómo se formarían los grupos experimental y de control. 3. Especificar la forma de hacer incidir el estímulo sobre el grupo experimental y la forma de desarrollar el experimento. 5. Hacer su esquema e indicar su significación.

RESPUESTA

1. Unidad de observación: los estudiantes de sociología que empiezan la carrera. Variable dependiente o efecto a medir: Actitudes sociales de los estudiantes de Sociología. Variable independiente o causa: los estudios de Sociología.

2. Como grupo experimental habría que tomar los estudiantes de primer curso de Sociología de una Facultad o Escuela de Sociología. En este experimento no existe grupo de control, y por eso es un experimento imperfecto. No es posible formar un grupo de estudiantes de sociología de primer curso no sujeto al estímulo de estudiar sociología, porque entonces dejarían de ser estudiantes de sociología.

3. El estímulo actúa de por sí, por el hecho de estudiar primer curso de esta carrera; por ello no hay que introducirlo. Lo único necesario es medir su efecto al comienzo y en las etapas sucesivas del curso. Para esto, como se trata de comprobar los cambios en las actitudes sociales derivados de las enseñanzas sociológicas, sería preciso construir un cuestionario y mejor escala, si ello es posible, para medir estas actitudes y opiniones, sobre todo las más en relación con las enseñanzas sociológicas impartidas en primer curso.

Al principio se aplicaría la escala o cuestionario, y después se iría repitiendo su aplicación a lo largo del curso, por ejemplo, cada tres meses.

Al ser instrumentos aplicados pasados períodos de tiempo apreciables, habría que contar con la posibilidad de la influencia de otros factores en los cambios de opinión detectados, por lo cual se incluirían en los cuestionarios preguntas generales en relación a los factores principales distintos del estudiado, que pueden influir en el cambio de aptitud y opinión, para ver su transformación en el período examinado y tenerlo en cuenta en el análisis de los resultados.

En cuanto al tipo de experimento se trata, como se ha indicado, de un experimento imperfecto o cuasiexperimento.

El modelo sería el siguiente:

	Actitudes sociales al comienzo de curso	*Actitudes sociales al final*
Grupo experimental	X_1	X_2
Grupo de control		

Variación en las actitudes: $d = X_2 - X_1$

Significación: Si la diferencia "d" es significativamente alta, indica que se puede seguir manteniendo la hipótesis de que los estudios de sociología influyen en las actitudes sociales si no hay indicación de que hayan actuado otros factores significativos. Pero si no hay diferencia o es escasa, entonces el experimento no confirma la hipótesis.

81

Supuesto que se tiene que realizar un experimento para comprobar la hipótesis de la influencia del alcoholismo en la avenencia conyugal y en general en la integración familiar, se pide: 1. Analizar la hipótesis y señalar la variable que ha de servir de estímulo en el experimento. 2. Señalar cómo se podrían formar el grupo experimental y de control, mediante el procedimiento de control de distribución de frecuencias. 3. Especificar la forma de hacer incidir el estímulo sobre el grupo experimental y la manera de desarrollar el experimento. 4. Indicar el tipo de experimento, hacer su esquema y explicar su significación.

RESPUESTA

1. Análisis de la hipótesis.

Unidad de observación: la familia.

Variable dependiente: la avenencia conyugal y en general la integración familiar.

Variable independiente: el alcoholismo.

Estímulo: la variable independiente, el alcoholismo.

2. Para formar los grupos experimental y de control, por el procedimiento indicado de distribución de frecuencias, se debería elegir dos grupos de familias que, en su término medio y en sus desviaciones de este término medio, fueran similares en características tales como género de ocupación, grado de instrucción, religiosidad, edad, origen social, situación económica, etc., diferenciándose en que en uno el cabeza de familia es alcohólico y en otro no.

3. El estímulo no se puede hacer incidir en el grupo experimental, por ser condición del enunciado que ya esté presente en él al formar éste. Para medir la incidencia del estímulo en el grupo experimental se aplicarían entrevistas, cuestionarios y escalas a ambos grupos que pusieran de manifiesto el grado de desorganización y de falta de avenencia familiares.

4. El tipo de experimento no es proyectado, sino natural, *ex-post-facto*, en cuanto el estímulo ha actuado ya antes de planear el experimento. Dentro de la categoría *ex-post-facto*, se trata de un experimento de causa a efecto, porque se quiere medir el efecto de la desorganización familiar y la falta de avenencia conyugal, partiendo de la causa, el alcoholismo. En el modelo de este experimento no hay medida en la fase inicial de los grupos experimental y de control. Su esquema sería por tanto, el siguiente:

	Grado de ajuste conyugal y organización familiar en la fase final del experimento	*Grado de ajuste conyugal y organización familiar en la fase inicial del experimento*
Grupo experimental . . .	X	
Grupo de control . . .	X'	

Variación del grado de ajuste y organización en la fase final, "d"

$$d = X - X'$$

Significación: Si la diferencia "d" es relativamente alta, indica que se puede seguir manteniendo la hipótesis de la influencia del alcohol en la avenencia conyugal y la integración familiar. Pero si no existe diferencia o es reducida hay que rechazar la hipótesis.

82

Supuesto que se pretende realizar un experimento para comprobar la hipótesis según la cual, dentro de las familias con varios hijos, los hijos o hijas primogénitos suelen ser de carácter más autoritario que los no primogénitos, se pide: 1. Analizar la hipótesis y señalar la variable que ha de servir de estímulo en el experimento. 2. Señalar cómo se podrían formar el grupo experimental y el de control mediante el procedimiento de control de distribución de frecuencias. 3. Especificar la forma de hacer indicar el estímulo sobre el grupo experimental y la manera de desarrollar el ejercicio. 4. Indicar el tipo de experimento y el esquema que correspondería a este tipo.

RESPUESTA

1. Análisis de la hipótesis.

Unidad de observación: Hijos de familia no únicos.

Variable dependiente o efecto: Carácter más tradicional, conservado y autoritario.

Variable independiente o causa: Hijos primogénitos o no.

2. Es preciso formar los dos grupos a contrastar, el de primogénitos y no primogénitos con muchachos o muchachas de edad similar que vivan en el mismo ambiente, por ejemplo un Colegio Mayor, y tengan un grado similar de educación. También es necesario que las condiciones sociales de los padres por término medio y en la desviación de este término medio sean similares en ocupación, ingresos, raza, creencias, ideología, etc., considerados los grupos en su conjunto.

3. Constituidos los grupos, se aplicarían a los individuos de ambas escalas de actitudes y cuestionarios para medir su grado de autoritarismo.

4. El tipo de experimento es como el anterior *ex-post-facto*, porque el estímulo hijos primogénitos o no ya ha incidido antes de empezar el experimento, y de causa

a efecto, porque, partiendo de la causa hijos primogénitos o no, se quiere medir el efecto, carácter conservador u autoritario. El esquema y su significación es idéntico al del anterior ejercicio.

	Carácter autoritario en la fase final	*Carácter autoritario en la fase inicial*
Grupo experimental . . .	X	
Grupo de control . . .	X'	

Variación en el carácter autoritario de los hijos primogénito y no primogénitos: $d = X - X'$. Si la diferencia es alta, indica que se puede seguir manteniendo la hipótesis de que la primogenitura influye en el carácter autoritario. Pero si no hay diferencia o es reducida hay que rechazar la hipótesis.

16. OBSERVACION DOCUMENTAL

NOCION Y CLASES

Los hombres no sólo tienen sentidos que les permiten observar los fenómenos sociales en el momento mismo en que se producen, sino que con la escritura expresan por medio de signos estas observaciones con lo que logran, hasta cierto punto, que se materialicen y trasciendan su esencial fugacidad. En realidad, en las sociedades con escritura, donde cada vez tiene menos importancia y fiabilidad la transmisión oral, para que perdure toda observación es preciso que se recojan sus resultados por medio de la escritura o de otra forma documental.

Pero, además, en las sociedades modernas el registro de los acontecimientos sociales de todo tipo, con fines no científicos, sino informativos, se realiza de manera institucionalizada y permanente por medio de la prensa diaria o periódica, cuya finalidad es recoger y dar cuenta de los acontecimientos más diversos de la vida social.

Asimismo el hombre utiliza la escritura para consignar en diarios, memorias, autobiografías, de un lado, y en novelas, ensayos, obras científicas y técnicas, de otro, desde sus experiencias más íntimas hasta los conocimientos científicos y técnicos de una época, pasando por su visión del mundo y de la sociedad y su juicio sobre ellos.

Por otra parte, en las sociedades actuales, se lleva por organismos públicos y privados, cuenta ordenada y regular de los hechos más importantes de la vida social: demográficos, económicos, laborales, sanitarios, educativos, financieros, etc. Esta cuenta ordenada constituye la base de las estadísticas en las que se refleja y presenta de manera sistemática y regular la extensión cuantitativa de los fenómenos y procesos sociales más importantes.

Por si esto fuera poco, hoy las modernas técnicas capacitan al hombre para registrar y conservar los sonidos e imágenes de toda clase de sucesos incluído el contenido de cualquier documento, y permiten, mendiante las técnicas informáticas y sobre todo la red mundial de información, documentación y comunicación INTERNET, verlos y obtenerlos en el momento en que se desee y cualquiera que sea el lugar donde se encuentren.

Por último, incluso los objetos de toda clase, desde el más grandioso monumento hasta el más simple utensilio, tienen un valor documental en cuanto son restos producto de una sociedad determinada, cuya naturaleza, organización y nivel de desarrollo reflejan.

Todo este arsenal de escritos, películas, fotografías, reproducciones de sonidos y objetos de toda clase son documentos y constituyen el objeto de la observación documental. Esta se puede definir como aquel tipo de observación que versa sobre todas las realizaciones que dan cuenta de los acontecimientos sociales y las ideas humanas o son producto de la vida social y, por tanto, en cuanto registran o reflejan ésta, pueden ser utilizados para estudiarla indirectamente.

De la exposición acabada de realizar se deduce claramente la gran multiplicidad y diversidad de los documentos y la amplitud, por tanto, del campo que se ofrece a la observación documental. Todos los tipos mencionados constituyen en su conjunto un arsenal inmenso de fuentes para la investigación social prácticamente inagotables. En el se encuentran recogidas y reflejadas, desde tiempos muy remotos, si bien de manera dispersa, desordenada y fragmentaria, gran parte de las manifestaciones de la vida social de la humanidad en su conjunto y en cada uno de sus sectores.

De ahí la importancia de este tipo de investigación. Es insustituible en los estudios de carácter histórico sobre fenómenos sociales pasados. Estos necesariamente se han de basar en los documentos existentes de dicho tiempo. En cuanto a las investigaciones sobre la realidad social actual, además de constituir los documentos, especialmente las estadísticas, un complemento indispensable de los demás medios de observación de la realidad social, han de partir de las anteriores investigaciones realizadas de tipo similar, que en relación a ellas constituyen también fuentes documentales.

Etimológicamente, documento se deriva del latín "documentum", que significa enseñanza, modelo, prueba y éste de "docere", enseñar. De acuerdo con esta etimología, son documentos todas las realizaciones obra del hombre, en cuanto son indicio de su acción y nos pueden revelar sus ideas, opiniones y formas de actuar y de vivir.

Los tipos de documentos objetivamente se pueden clasificar según la exposición anterior:

a) En documentos escritos entre los que cabría distinguir como categorías principales: la prensa; los llamados documentos personales: cartas, diarios, memorias, autobiografías; las obras de creación literaria: novelas, obras de teatro, cuentos, ensayos, y todas las demás publicaciones impresas.

b) Los documentos numéricos o estadísticos de todo tipo.

c) Documentos de reproducción de la imagen y el sonido (cine, televisión, magnetófonos, discos).

d) Documentos-objeto: toda clase de realizaciones técnicas y artísticas del hombre.

Un tipo de documentos autobiográficos que merece mención especial es el llamado "biograma". Con él se designa las autobiografías escritas a petición de los investigadores.

FIABILIDAD DE LOS DOCUMENTOS ESCRITOS

Los documentos escritos, en general, como fuente de observaciones sociales, tienen carácter secundario, en cuanto no ofrecen los fenómenos sociales mismos que han tenido lugar, sino el resultado de la percepción e interpretación de ellos por el escritor, que siempre es parcial e incompleta. Por ello es obvio que presentan también un carácter indirecto, ya que no permiten al que los utiliza un contacto directo con los hechos, sino mediato, a través de los documentos.

Este carácter secundario o indirecto de los documentos, unido a que frecuentemente están producidos o escritos por no investigadores y para fines no científicos, exige como condición para su utilización en la investigación social un examen o juicio previos muy rigurosos de su fiabilidad.

En la fiabilidad de los documentos hay que tener en cuenta una distinción importante, por una parte entre los hechos y fenómenos que se describen y las interpretaciones que se dan de los mismos, y por otra, entre las ideas, opiniones y actitudes contenidas y expresadas al describir e interpretar los acontecimientos. Es obvio que la aceptación por válidas de las primeras, requiere mucho más cautela que la de las segundas.

Puntos básicos a tener en cuenta en el juicio crítico de la fiabilidad de los documentos son:

a) Su autenticidad.

b) Su alteración o no por interpolaciones o por lagunas.

c) La competencia del autor y su conocimiento directo o no del tema tratado.

d) Su marco teórico y su ideología.

e) La coyuntura política y social en que fue escrito.

f) La pertenencia del autor a algún partido político o grupo de presión.

g) El fin pretendido con la publicación del documento.

También hasta cierto punto, aunque en menor grado, sobre todo en algunos aspectos, estas circunstancias son aplicables a los documentos estadísticos. Además, en relación a este tipo de documentos existen las

reglas formuladas por Bowley, que se enumeran en el ejercicio 85 como base para el análisis de una estadística concreta. (Ver ejercicios 83 a 85)

Por otra parte, en la utilización de documentos no son de menor importancia los problemas anejos a la lengua en ellos utilizada y especialmente los referentes a su sentido exacto en la época en que fueron escritos.

ANALISIS PRIMARIO DE DOCUMENTOS

En el análisis de los documentos se pueden distinguir el análisis primario y el secundario, y dentro del primero el análisis pormenorizado del contenido de los documentos y su análisis como un todo global.

El pormenorizado se basa en el establecimiento previo de las variables empíricas y las categorías sobre las que necesitamos recoger información.

Una vez establecidas las variables y categorías se examinan sistemáticamente los documentos, con el objeto de encontrar los datos contenidos en ellos referentes a cada categoría. La finalidad pretendida es ver si los datos prueban o no las hipótesis formuladas.

En esto consiste, en sustancia, a mi juicio, la aplicación del método científico a la investigación documental en ciencias sociales, y a ello se dirige el análisis de contenido, que examinaremos después más detenidamente.

No obstante lo anterior, los documentos constituyen por si hechos sociales. Como tales pueden ser objeto, como un todo global, de diversos tipos de análisis sobre sus características, con intención de deducir de ellas consecuencias de interés social.

Entre estos tipos de análisis Duverger (1962, p. 152 y ss.), destaca y expone en relación a los documentos, los tipos de análisis interno, externo, psicológico y sociológico.

Se puede señalar como objetivo del análisis interno la determinación del mensaje de la obra y de los principios y valores en que se funda.

Por el contrario, parece objetivo más propio del análisis externo, además del examen de los aspectos formales, estilo, lenguaje, etc., la puesta en relación del documento con el contexto histórico y cultural en que tiene lugar su aparición.

Sociológicamente, se pueden analizar en el documento, los puntos, con un enfoque social, que señala Lazarsfeld. Quién habla: sus características sociales; para decir qué: mensaje social; a quién: grupos sociales a que se dirige; cómo: la forma del mensaje, y con qué resultados.

Desde un punto de vista psicológico se pueden estudiar la personalidad social de su autor y sus actitudes y opiniones en relación a cuestiones sociales.

Una modalidad especial de este tipo de análisis es el análisis de los documentos-objeto, consistente en estudiarlos en sus diversos aspectos y deducir de ellos las circunstancias sociales y económicas con las que pueden estar relacionados. Por ejemplo, Duverger (1962, p. 140), distingue en su estudio tres tipos de análisis: el material, que consiste en examinar externamente el objeto en sí mismo; el tecnológico, que atiende al uso práctico de los objetos e indica el nivel técnico que suponen. Por último, el análisis simbólico que se orienta, según él, a investigar los significados y valores que los hombres atribuyen a estos objetos. (Ver ejercicio 86).

ANALISIS DE CONTENIDO. ORIGEN, NOCION E IMPORTANCIA

El análisis de contenido es la técnica, sin duda más elaborada y que goza de mayor prestigio científico en el campo de la observación documental.

Su origen histórico es reciente. Su aplicación de una manera científica se puede decir que data de la II Guerra Mundial. Durante ella fué utilizada por los aliados, de manera regular, con éxitos notables, para el estudio de las informaciones dadas en los medios de comunicación nazis.

Asimismo, Lasswell utilizó ésta técnica para el estudio de la propaganda bélica en la II Guerra Mundial. Este autor en 1949 publicó la obra *"Language of politics. Studies in quantitative semantic."* Junto a la de Berelson, *"Content analysis in comunication researchs"* de 1952 son las obras clásicas de esta técnica.

Desde 1970, la evolución del análisis de contenido (Bardin, 1973) se distingue por tres rasgos: 1. El recurso al ordenador. 2. El interés por las comunicaciones no verbales y la semiología. 3. Su confluencia con el gran desarrollo actual de la lingüística.

El autor indicado, Bernard Berelson, la define de la siguiente manera:

"Es una técnica de investigación para la descripción objetiva, sistemática y cuantitativa del contenido manifiesto de las comunicaciones, con el fin de interpretarlas."

Según esta definición, las notas que caracterizan al análisis de contenido son éstas: ser objetivo, sistemático, cuantitativo, sobre el contenido manifiesto y con el fin de interpretarlas.

La primer nota, objetivo, supone el empleo de procedimientos de análisis que puedan ser reproducidos por otros investigadores, de modo que los resultados obtenidos sean siempre susceptibles de verificación.

La segunda nota, sistemático, exige la sujeción del análisis a una pauta objetiva determinada, que abarque todo el contenido a observar.

La tercera nota, cuantitativo, pide que se puedan cifrar numéricamente los resultados del análisis. En relación a esta nota hay que tener en cuenta que, como describe D.P. Forcese (1973, p. 186), el análisis de contenido es una técnica para cuantificar datos secundarios: consiste básicamente en el aislamiento y recuento de unidades e indicadores de los fenómenos en que estamos interesados.

Por ello, el análisis de contenido es, como afirma Bailey (1978, p. 276), equivalente en el estudio de documentos a la investigación por encuesta y sus resultados se condensan, como en ésta, en tablas numéricas.

La cuarta nota, manifiesto, indica que el objeto de éste análisis son los contenidos expresos de la comunicación, en el texto de que se trate. Objeto inmediato de observación, sólo pueden ser los contenidos manifiestos. De hecho, sólo éstos se pueden observar directamente. Pero esto no impide que se pretenda también investigar contenidos latentes, utilizando para ello indicadores manifiestos que sean reflejo o expresión de los mismos.

Esta conclusión resulta implicada en la quinta nota: con el fin de interpretarlas. Lo que el análisis de contenido pretende no es quedarse con el contenido manifiesto, sino inferir de él consecuencias relevantes de naturaleza psicológica, sociológica, política, histórica, etc. sobre el origen, destino y aspectos de los mensajes.

El objeto del análisis de contenido consiste concretamente en observar y reconocer el significado de los elementos que forman los documentos (palabras, frases, etc.) y en clasificarlos adecuadamente para su análisis y explicación posterior.

En las sociedades desarrolladas y masificadas modernas, la escritura y los medios de comunicación ocupan una posición cada vez más destacada y predominante.

En estos medios se recogen y reflejan la vida de las sociedades modernas y los valores, patrones culturales y actitudes ante los problemas del hombre y la sociedad, vigentes en cada momento.

Por ello su contenido constituye un material indispensable para el conocimiento de estas sociedades.

En consecuencia, es clara la importancia del análisis de contenido, en cuanto técnica para el análisis científico de todos estos materiales.

FASES DEL ANALISIS DE CONTENIDO

El análisis de contenido no es otra cosa que una modalidad especial de la aplicación del método científico en las ciencias sociales. Sus operaciones son por tanto, fundamentalmente, las mismas indicadas al hablar de la investigación social en general.

Por ello me limitaré a ocuparme de las que por el objeto de análisis de contenido, presentan particularidades más destacadas: el muestreo, la determinación de unidades de análisis, la elección de categorías y la confección del cuadro de recogida de datos.

La utilización de una muestra en el análisis de contenido, pueden ser muchas veces indispensables en la práctica, dada la multiplicidad frecuente de unidades documentales a observar. La particularidad que presenta el muestreo en este caso es que aquí no se refiere a personas, sino a fuentes documentales, números o espacios de periódicos y de revistas, ejemplares de libros, emisiones de radio y TV, etc.

Los tipos principales de muestreo en este caso son tres que no se excluyen entre sí: de fuentes, por ejemplo, elección entre los distintos títulos de periódicos y revistas; de fechas, muestreo entre los números de cada fuente; y de espacios: dentro de un mismo documento, se pueden distinguir espacios iguales, numerarlos y sortear entre ellos.

Merece mención especial el muestreo de números o el de emisiones. En él hay que tener en cuenta que , dado el carácter periódico de estas fuentes documentales, la base de la muestra que se elija o forme, generalmente no será homogénea, y presentará diferencias según los meses, años, estaciones, semanas y los días de la semana.

Una forma de obviar este inconveniente es adoptar un sistema de muestreo rotatorio, en el que se vayan cambiando, por ejemplo, los días de la semana en los sucesivos meses.

Unidades de análisis.— Las unidades de análisis en el análisis de contenido son modalidades especiales y particularizadas de las unidades de observación.

Si en un análisis de contenido que verse sobre la prensa se puede decir que la unidad de observación genérica es cada número del diario, las unidades de análisis concretas dentro de cada número del periódico pueden ser muy diversas, según el objeto de la investigación, por ejemplo:

las palabras, las frases, los párrafos, los artículos, los temas tratados, las secciones del periódico, los títulos de los artículos, etc.

Modalidades fundamentales de estas unidades son las unidades de contexto, las de registro y las reglas de enumeración. Las unidades de contexto están formadas por las partes o segmentos del contenido de los documentos necesarias para comprender el significado de las unidades de registro. Puede ser el ámbito, espacio de texto entre dos puntos y seguido consecutivos, el párrafo, el parágrafo, el capítulo, el documento entero.

La unidad de registro es la unidad de significación que se busca en el texto, se extrae del mismo y se codifica. Sus modalidades pueden ser también muy diversas, tales como palabras, temas o afirmaciones sobre algo, asuntos generales tratados, personajes que intervienen en el mensaje, los sucesos o acontecimientos que se narran en él, o el documento mismo como un todo.

Las unidades de registro se cuentan en el análisis de contenido pero las maneras de contarlas pueden ser muy diversas. De aquí la necesidad de reglas de enumeración que precisan dicha forma de recuento. Entre ellas cabe destacar, como más común de todas, la frecuencia o número de veces que se repite la unidad de registro en la unidad de contexto; la simple presencia o ausencia de la unidad de registro; la gradación de la intensidad de significación de la comunicación; el orden de aparición de las unidades de registro; la co-ocurrencia o presencia simultánea de dichas unidades.

Uno de los análisis de contenido más simple y empleado, es el que versa sobre los temas o asuntos. Por ejemplo, es frecuente analizar los contenidos de la prensa, emisiones de radio y TV, clasificando este contenido según la importancia concedida y la extensión que ocupan en estos medios informativos, cada clase de asuntos tratados: políticos, religiosos, educativos, artísticos, musicales, deportivos, etc. En este caso las unidades de análisis suelen ser el artículo, la noticia, el comentario, etc., completos.

También es corriente establecer una lista de palabras clave en una ideología, y sus contrarias como base de referencia y comparación, y luego contar las frecuencias con que se encuentran en los documentos estudiados. Aquí las unidades de análisis son las palabras.

Un tipo de análisis intermedio entre éste y el anterior es establecer un conjunto de proposiciones características de una actitud, doctrina, ideología, partido, etc y revisar las ideas expuestas en el documento expresando numéricamente la frecuencia de acuerdo o desacuerdo con ellas. En este caso, las unidades de análisis son generalmente los párrafos.

Para terminar, se puede igualmente analizar un texto fijándose en el tipo de aserciones empleadas por el autor preferentemente: activas o pasivas, referentes al pasado, al presente o al futuro; relativas al autor o a otras personas. En este caso las unidades de análisis son las oraciones.

ELECCION DE CATEGORIAS

En la investigación social, se llaman categorías a cada uno de los elementos o dimensiones que comprende una variable cualitativa. Por ejemplo, la variable profesión, comprende las categorías obrero, empleado administrativo, funcionario civil, militar, etc.

Las categorías representan unos elementos más concretos, definidos y singulares que las variables empíricas o indicadores. Constituyen cada uno de los elementos singulares que vamos a buscar en la investigación, referentes a determinadas variables.

De acuerdo con esta noción, en el análisis de contenido se llaman categorías a cada uno de los elementos o dimensiones últimos y más simples de las variables investigadas, que nos van a servir para clasificar o agrupar según ellas las unidades de análisis del texto.

Es obvio que no se puede hacer una enumeración de categorías posibles en el análisis de contenido. Por eso sólo aludiremos por ejemplo a algunos de los tipos más frecuentes.

Categorías de materia o asuntos. Se refieren a los temas tratados por los documentos, por ejemplo: políticos, económicos, sociológicos, laborales, religiosos, etc.

Categorías referentes al sentido de la comunicación en relación a alguna cuestión: Ejemplo de estas categorías lo son las de favorable, desfavorable, neutral, etc.

Las categorías de valores, o de lo que la gente desea, quiere o busca. Son categorías de este tipo, por ejemplo, el dinero, la salud, el amor, el éxito, etc.

Categorías sobre formas de acción posible. Entre ellas se encuentran, el análisis de los hechos, la propaganda, la negociación, la organización, la utilización de medios económicos, la violencia, etc.

Categorías referentes a los actores. Aquí se engloban todas las categorías referentes a las variables básicas de identificación de los sujetos humanos. Son las referentes al sexo: hombre y mujer; al estado civil: casado, viudo, soltero, etc.

Categorías que indican una toma de posición del autor, tales como la aprobación, desaprobación, pesimismo, optimismo, afirmación, negación.

Las categorías como los indicadores, constituyen la red utilizada en la investigación. Por ello de la perfección o imperfección del conjunto que se forme depende el éxito o fracaso del trabajo emprendido.

Para el establecimiento de categorías hay que tener en cuenta el objetivo de la investigación y el material a investigar. En relación a éste se recomienda releerlo varias veces. Esta lectura facilitará la determinación de las categorías a emplear.

En todo caso se debe probar con ensayos previos la validez y seguridad de las categorías elegidas.

El cuadro.– El instrumento de observación básico de que se dispone hasta ahora en el análisis de contenido es el cuadro.

El cuadro deberá comprender todas las categorías sobre las que se va a recoger información en los documentos.

Cada categoría ha de tener los espacios necesarios para anotar en ellos la intensidad, frecuencia y las características que se quieren registrar sobre cada una de ellas.

A cada unidad de observación o documentos independientes se dedicará un cuadro, o más si son necesarios, en el que se consignarán los datos que identifiquen dicha unidad.

Si no se está seguro dónde clasificar un contenido, se puede tomar nota de él y después decidir a que categoría antigua o nueva se atribuye.

Se recomienda anotar en cada contenido clasificado la categoría en que lo fué, para que se pueda contrastar la exactitud y precisión de la clasificación.

Se deberá también contrastar la seguridad del cuadro, mediante la aplicación al mismo material por distintos observadores, así como su validez. (Ver ejercicios 87, 88, 89)

ANALISIS SECUNDARIO DE DOCUMENTOS

Se entiende por análisis secundario de documentos o de datos documentales (Hakim, 1982) todo posterior análisis de un conjunto de datos primarios que ofrezca interpretaciones y conclusiones adicionales o en forma diferente a la presentada en el primer informe de investigación.

Según esta definición, el análisis secundario de datos implica dos requisitos:

1. Un análisis posterior o reanálisis de datos obtenidos y analizados anteriormente.
2. Que el nuevo análisis no sea una repetición de los análisis conocidos ya realizados con los datos, sino que, por abarcar más datos que éstos, basarse en tratamientos diferentes de los datos, emplear tipos de análisis diversos, ofrezca interpretaciones y conclusiones adicionales o en forma diferente a las ofrecidas en el primer análisis.

Las fuentes principales de los datos que se emplean en el análisis de datos secundarios numéricos son:

a) Las estadísticas de todo tipo.

b) Los protocolos que contienen el conjunto de resultados numéricos en los que se concretan las investigaciones científicas realizadas, mediante encuestas, experimentos, etc.

c) Los bancos de datos.

Dentro de las estadísticas, merecen mención especial las realizadas por los Institutos Nacionales de Estadística. Estos organismos, encargados primero de realizar los censos de población, suelen tener a su cargo en la actualidad la realización periódica de grandes encuestas nacionales, por ejemplo, en España, sobre el empleo y los presupuestos familiares.

En cuanto a los protocolos de investigaciones, en ellos se contienen las medidas de las variables investigadas obtenidas en la observación y en general los análisis numéricos efectuados y sus resultados. Esta masa de datos no es explotada o analizada exhaustivamente en las investigaciones originales y, por tanto, puede ser objeto de nuevos análisis secundarios.

Los bancos de datos se caracterizan por la aplicación de medios informáticos para el almacenamiento ordenado y la recuperación de datos numéricos, que pueden proceder, bien de fuentes estadísticas, o de protocolos de investigaciones realizadas. Por tanto, se trata de una fuente complementaria de las otras dos indicadas.

El análisis de datos secundarios puede tener múltiples aplicaciones, entre las que se pueden destacar las siguientes:

a) La elaboración de informes tanto generales como particulares o referentes a un tema muy específico.

b) Realización de análisis basados en un nuevo marco teórico distinto de los ya empleados anteriormente en los datos.

c) Reanálisis mediante la aplicación de nuevas técnicas y para contrastar nuevas hipótesis.

Sobre todo, se debe subrayar que también se pueden realizar investigaciones importantes mediante la utilización de datos viejos o secundarios y que la originalidad de una investigación no depende de que en ella se utilicen datos primarios.

Entre sus ventajas, señaladas por los autores (Hackim, 1982), se pueden destacar las siguientes:

1. Es posible que el investigador creativo pero desorganizado no trabaje a gusto en el equipo de una gran encuesta; si se especializa en el análisis secundario, puede trabajar solo y utilizar los datos recogidos por un equipo bien organizado.
2. Los investigadores, que buscan evidencia empírica para teorías, pueden evitar la nueva colección de datos valiéndose del análisis secundario.
3. Los investigadores interesados en el desarrollo de indicadores sociales se deberán basar exclusivamente en datos estadísticos.
4. En el caso de falta de recursos económicos, la investigación académica, tesis doctoral por ejemplo, basada en datos secundarios puede ser una alternativa al pequeño estudio local o a la larga disertación teórica, dado que normalmente es mucho más barata que la investigación empírica exigida por una nueva colección de datos.

Por último, cabe subrayar por Boudon (1975), la necesidad para el desarrollo de la investigación social y de las ciencias sociales de la formación de una contabilidad social y el establecimiento de bancos de datos sociales.

EJERCICIOS

83

Supuesto que se tienen que realizar estudios basados en los documentos que se indican, se pide exponer razonadamente el grado de fiabilidad que merecen como fuentes documentales.

1. Un periódico de un partido político en lo referente a problemas políticos.

RESPUESTA

Los periódicos de partidos políticos son poco fiables en cuestiones políticas, pues tienden, según su ideología, a deformar los hechos, a presentarlos de manera parcial e incluso omitir los que le perjudican. Es preciso, pues, contrastarlos con otros de distinto signo y, si es posible, con fuentes distintas.

2. Un folleto en que se enumeran para conocimiento general las realizaciones sociales de una gran empresa.

RESPUESTA

Ni por su origen ni por su destino es muy fiable este documento, ya que procede de parte interesada en destacar dichas realizaciones sociales y no tiene una finalidad técnica, sino publicitaria y vulgarizadora. La utilización de sus informaciones exige algún género de contraste.

3. Un libro del siglo XVII escrito por un profano en el que se exponen los adelantos en Medicina de su tiempo.

RESPUESTA

Tampoco es fiable esta fuente, y no se puede utilizar sin verificación, aparte de otros motivos, al proceder de persona profana o no competente en el campo médico.

4. Una revista dependiente de un importante grupo financiero con relación a hechos y cuestiones económico-sociales discutidos.

RESPUESTA

Como en el caso número 1, hay que poner entre paréntesis la fiabilidad de esta revista sobre dichos hechos y cuestiones discutidos. Es de suponer que el grupo de presión financiero imponga su línea ideológica a la revista según la cual se juzguen los hechos y cuestiones.

5. Una obra del siglo XVII en la que se describen las costumbres de los indios americanos, sin haber estado allí el autor, basándose en el relato de otras personas.

RESPUESTA

Esta obra tampoco es fiable por describir hechos no conocidos directamente por el autor y con base en una fuente secundaria de tan poca garantía como los relatos.

6. La descripción de un viaje a la China maoísta, realizada diez años después del viaje, sin que se tomaran entonces notas previas.

RESPUESTA

También es dudosa la fiabilidad de esta descripción, en cuanto está hecho después de pasado mucho tiempo del viaje y con base sólo en la memoria.

7. Una publicación sobre las causas de la segunda guerra mundial, escritas por un alemán nazi en plena guerra mundial.

RESPUESTA

No es fiable por proceder de uno de los contendientes y estar escrita en la misma guerra sin tener, por tanto, perspectiva histórica y hallándose inmerso en el ambiente social belicista.

8. Un diario o memorias de un autor con relación: a) a hechos discutidos en que tomó parte; b) a cuestiones indiferentes para el autor; c) a las declaraciones perjudiciales para él, y d) juicio de sus adversarios.

RESPUESTA

Los diarios, memorias y documentos personales son menos fiables en lo referente a cuestiones personales del autor y sus adversarios —nadie es buen juez en su propia causa—, y más en las cuestiones indiferentes, las perjudiciales para él y las públicas, porque las primeras no afectan a sus intereses, porque es poco probable en la segunda que, sin creer que son verdad, alguien se perjudique con sus declaraciones, y es difícil, respecto a las terceras, que se pretenda negar lo conocido por todos.

84

Se está realizando un estudio sociológico que se basa en documentos estadísticos.

Se pide analizar la fiabilidad de las fuentes estadísticas que se indican:

A) Estadísticas fiscales basadas en las declaraciones de los contribuyentes respecto a las actividades económicas gravadas.

RESPUESTA

No son fiables en absoluto, pues se basan en las declaraciones de los sujetos interesados, y éstos tienen interés en declarar lo menos posible para tributar poco. De hecho, es conocida la extensa defraudación en este campo.

b) Una estadística en la que se compara la variación del volumen e importe de las exportaciones en los años 1964 al 1973, importe que se expresa en pesetas de cada año.

RESPUESTA

Tampoco es válida la comparación de esta estadística, en cuanto compara cosas distintas, pesetas de cada año de 1964 a 1973. Lo que habría que hacer para que la comparación fuera válida es deflactar las pesetas de cada año, es decir, reducirlas a pesetas de 1964, teniendo en cuenta el índice de variación en el poder adquisitivo de nuestra moneda durante dichos años.

c) Una estadística tomada de un trabajo publicado en una revista en el que se cita como fuente de donde la ha tomado el autor el "Boletín Estadístico del Banco de España".

RESPUESTA

En este caso se trata de una fuente secundaria. Por eso lo que procede, si se piensa utilizar dicha estadística en una investigación, es compulsar con la fuente de origen, *Boletín Estadístico del Banco de España*, y tomarlas de esta publicación.

d) Una estadística publicada en un periódico sin indicar fuente, sobre la distribución personal de la renta en España según clases sociales.

RESPUESTA

Esta estadística, al ser anónima, no ofrece garantía alguna, pues no se puede comprobar con la fuente de donde procede, ni juzgar la fiabilidad de esta fuente y de los procedimientos empleados para obtenerla.

e) Una estadística sobre la natalidad rural y urbana en España que se basa en los datos del Registro Civil.

RESPUESTA

No es enteramente fiable esta estadística sobre natalidad, en cuanto está viciada a los efectos pretendidos, por el hecho de que en la actualidad muchas madres que viven en el campo van a dar a luz a clínicas de la ciudad, donde son inscritos sus hijos en el Registro Civil. En consecuencia, las estadísticas basadas en datos de dicho Registro la computan como natalidad urbana, a pesar de ser de hecho campesina.

f) Las estadísticas sobre turismo en Italia formadas por los organismos públicos encargados del turismo.

RESPUESTA

No se debe conceder fiabilidad absoluta a esta estadística, pues el organismo de que proceden tiene interés en que aparezcan lo más elevadas posible, y sin duda procura de hecho inflarlas algo.

g) Un folleto editado por Alemania Occidental con propósitos divulgadores entre el público general sobre los logros económicos de la República Federal.

RESPUESTA

Tampoco es fiable esta fuente. Primero, por ser secundaria y, segundo, por su destino publicitario entre el público en general.

85

Dada la siguiente estadística, tomada del "Anuario Estadístico de España, edición manual de 1971, se pide analizarla críticamente, siguiendo las reglas de Bowley que se transcriben a continuación:

1. Examinar la definición de las unidades clasificadas y el procedimiento con el que se obtuvieron los datos o tratar de determinarlos cuando no se consignan como es frecuente.

2. MOVIMIENTO NATURAL DE LA POBLACION

2.1. Matrimonios, nacimientos y defunciones

2.1.1. Datos generales

CONCEPTO	1966	1967	1968	1969	1970 (3)
Cifras absolutas					
Matrimonios	222.796	232.624	231.546	238,973	246.592
Nacimientos (1)	661.731	672.039	659.677	658.931	655.495
Defunciones (2)	269.738	274.021	277.357 (3)	297.125	281.777
Crecimiento vegetativo	391.993	398.018	382.320 (3)	361.805	373.718
Por mil habitantes (4)					
Matrimonios	7,19	7,20	7,10	7,25	7,41
Nacimientos (1)	20,70	20,81	20,22	20,00	19,69
Defunciones (2)	8,44	8,49	8,50 (3)	9,02	8,46
Crecimiento vegetativo	12,26	12,33	11,72 (3)	10,98	11,23

(1) Nacidos vivos.– (2) Sin los nacidos muertos, muertos al nacer o antes de las veinticuatro primeras horas de vida.– (3) Cifras provisionales.– (4) Según la población calculada para el primero de julio de cada año.

Anuario Estadístico de España. Ed. Manual, 1971.

2. Examinar si las cosas, objetos, etc., que se presentan agrupados son semejantes realmente.

3. Examinar si existe alguna correlación entre las cantidades medidas por el censo y la que deseamos conocer, teniendo en cuenta la posible influencia de otras variables.

4. Cuando se trabaje con promedios, tasas o tantos por cientos examinar bien las unidades del numerador y denominador, si son homogéneos y si se relacionan bien.

5. Cuando se comparan dos cantidades comprobar si en realidad expresan unidades comparables en la observación.

6. Investigación de la precisión o errores en el método de recolección de datos.

7. No fiarse de las cifras de un año, un mes o un día, sino establecer comparaciones sobre datos devarios años, meses, etc.

8. Examinar rígidamente las conclusiones deducidas de los datos, para ver si se derivan realmente de ellos o no se hace decir a los datos lo que el investigador quiere.

RESPUESTA

1. En este caso únicamente se da la definición de nacimientos y defunciones. Según ella sólo se toman en cuenta los nacidos vivos, no considerándose como tales los muertos al nacer o dentro de las veinticuatro primeras horas de vida. No da la definición de matrimonio, pero se ha de estimar que considera como tal todos los inscritos en el Registro Civil, religiosos y civiles. En cuanto al crecimiento vegetativo, se ve que es la diferencia entre los nacimientos y defunciones computados. Tampoco se dice nada sobre el procedimiento de obtención de los datos.

2. No hay ningún motivo para dudar que los nacimientos, defunciones y matrimonios que se presentan agrupados no sean semejantes realmente.

3. En este aspecto, la estadística es válida siempre que se quieran estudiar dichos fenómenos demográficos con referencia a España.

4. Las tasas que se dan han sido obtenidos de las cifras absolutas de esta misma tabla, puestas en relación con la población calculada para el primero de julio de cada año. Todo ello es correcto, pues se relacionan unidades homogéneas, personas. En cuanto a la población calculada en primero de julio de cada año debe considerarse bastante fiable, al realizarse por el organismo técnico competente en la materia el Instituto Nacional de Estadística.

5. En esta tabla existe una comparación temporal, años 1966 a 1970, y una comparación por conceptos, matrimonios y defunciones, en cifras absolutas y porcentajes. Es claro que al ser en todos los casos las unidades las mismas e invariables con el tiempo, son comparables. Otro caso sería, por ejemplo, si se compararan pesetas de los distintos años sin tener en cuenta la pérdida de valor de la moneda.

6. Los datos, aunque no se indica, deben proceder de las inscripciones en el Registro Civil, elaborados después estadísticamente por el Instituto Nacional de Estadística. Aunque no se puede descartar en absoluto la existencia de errores, bien de inscripción en el registro o de cómputo, éstos son, sin duda, muy raros, y se deben tener como no significativos.

86

Se está realizando una investigación social en una pequeña comunidad rural, muy pobre, dedicada principalmente al pastoreo, y entre los documentos-objeto hallados se encuentra el empleo del arado de madera.

Analizar sociológicamente este documento según sus aspectos material, tecnológico y simbólico.

RESPUESTA

Análisis material: al estar hecho el arado de madera, muestra una gran pobreza y el aislamiento social de dicha comunidad, en relación al mundo moderno actual.

El examen tecnológico indica un nivel técnico muy elemental del desarrollo de la agricultura, indicio a su vez de una organización económica y social rudimentarias y todavía primitivas.

Por último, del análisis simbólico se puede derivar que, dentro de la comunidad social en cuestión, el arado de madera puede ser un signo de prestigio social, como elemento de diferenciación del resto de la comunidad, que se dedica, según se ha señalado, al pastoreo.

87

Supuesto que se piensa realizar un estudio de contenido del tipo y sentido de las informaciones en 1975 de la prensa extranjera sobre cuestiones sociales, políticas y económicas españolas, se pide:

1. Formular la hipótesis o hipótesis que parezcan probables sobre el tipo y sentido de las informaciones en cuestión.

2. Analizar estas hipótesis.

3. Establecer las categorías.

4. Delimitar el campo de investigación.

5. Determinar el procedimiento de muestreo.

6. Establecer el procedimiento de recogida de datos y señalar los instrumentos a emplear para recogerlos.

RESPUESTA

1. Aunque se trata de un estudio descriptivo, se puede formular, por ejemplo, las hipótesis de trabajo siguientes:

a) En las informaciones de la prensa extranjera sobre España es probable que predominen las de tipo político.

b) En las informaciones de la prensa extranjera sobre España es probable que predominen las de signo desfavorable.

2. Análisis de las hipótesis. En las dos, la unidad de observación es la misma, la prensa extranjera.

Solamente tienen ambas una variable: el tipo de informaciones sobre España, la primera, y el signo de dichas informaciones, la segunda.

3. Establecimiento de categorías. Para efectuar esta operación sería conveniente realizar un estudio previo de la prensa extranjera sobre España tomando notas de los temas que suele tratar para establecer las categorías de acuerdo con este resultado. En cuanto a las categorías del signo de la información, se podrían distinguir dos posiciones fundamentales, favorable y desfavorable, y dentro de ellas dos o tres categorías intermedias según la intensidad de la posición en favor o en contra.

En un estudio realizado por el Instituto Español de la Opinión Pública publicado en la *Revista de Opinión Pública*, número 20, de 1970, se establecieron las siguientes categorías generales referentes a temas tratados: 1. Información económica y financiera.— 2. Información nacional.— 3. Información sobre relaciones internacionales.— 4. Información laboral y social.— 5. Información sobre la Iglesia española.— 6. Información cultural e histórica.— 7. Monarquía.— 8. Turismo. Según su signo, se clasificaron en: muy favorable, desfavorable, fluctuante, desfavorable, muy desfavorable y neutra.

4. Delimitación del campo de investigación. Temporal: Podría abarcar un año, por ejemplo, el 1975, como se dice en el enunciado. También se podría extender además a otro año de otro período anterior, a fin de establecer comparaciones.

Geográfica: Habría que determinar los países a comprender en el trabajo. Limitaciones al efecto serían la recepción de manera regular de prensa, así como la disposición de especialistas en la lengua de todos ellos. Dentro de los países posibles la elección se haría por decisión razonada, tendiendo a escoger los países más representativos de las diversas áreas geográficas y tipos de países.

Definición del universo: En primer lugar, sería preciso especificar el significado de prensa, por ejemplo, sólo diarios de información general, lo que significaría prescindir de todo otro tipo de publicación periódica: semanal, quincenal, etc., así como de cualquier posible tipo de publicación diaria especializada.

5. Como sería demasiado trabajo consultar todos los números de cada periódico, se debería formar una muestra, tomando, por ejemplo, de cada título un día a la semana. El primero se elegiría por sorteo, y luego se podría establecer un sistema rotatorio cambiando correlativamente cada mes el día de la semana elegido.

6. Procedimiento e instrumento de recogida de datos. Se podrían utilizar fichas en las que en la cabecera se indicaría el nombre del periódico, su nacionalidad, orientación ideológica, número, fecha, y en el cuerpo se dedicarían las filas a la categoría elegida del tipo de noticias, y las columnas, a la extensión, frecuencia y categorías referentes al sentido de la noticia.

88

La enseñanza universitaria pública está estructurada y persigue determinados objetivos en función de ciertos valores. Esto supuesto, se puede uno plantear el problema de si estos valores son permanentes o varían con el tiempo y las transformaciones sociales.

En relación al problema acabado de plantear se piensa estudiar la permanencia o evolución de los valores a que se orienta la enseñanza universitaria en España con base en las actas de las sesiones de apertura de curso de las universidades españolas.

Diseñar con este fin un plan de investigación documental que comprenda:

1. Formulación de la hipótesis de trabajo.

2. Su análisis.

3. Determinación de categorías de las variables.

4. Delimitación del campo de investigación.

5. Determinación del procedimiento a seguir para la recogida de datos y confección del instrumento oportuno al efecto.

RESPUESTA

1. Formulación de la hipótesis: Dado que en general parece detectarse una transformación de los valores sociales orientada hacia el predominio de los valores económicos, utilitarios y prácticos sobre los tradicionales y espirituales, se podría formular la siguiente hipótesis.

Los fines y valores perseguidos por la enseñanza universitaria en España, según los actos de apertura de curso de las universidades, deben reflejar probablemente una transformación en el sentido de un predominio progresivo de los fines y valores económicos y prácticos sobre los tradicionales e ideales.

2. Análisis de la hipótesis. Unidad de observación: Las actas de apertura de curso de las universidades. Variables: No hay variable independiente porque no es hipótesis causal. Variables dependientes o efectos estudiados: 1. Valores económicos y prácticos. 2. Valores ideales y tradicionales.

3. Determinación de categorías en las variables: Para establecerlas sería conveniente realizar un sondeo previo en las actas de apertura de curso. A título de ejemplo, se indican las siguientes categorías para los valores ideales: Valores religiosos y morales, valores patrióticos, valores culturales humanísticos, valores tradicionales, valores de integración social, valores de refinamiento individual buscado por sí mismo. Para los valores utilitarios: Valores de adquisición de técnicas y habilidades operativas, valores de adquisición de conocimientos para transformar el mundo exterior, valores de elevación del bienestar material, valores del rendimiento económico, valores de aumento de la productividad, valores de contribución al desarrollo económico del país.

4. Delimitación del campo de investigación: En principio, todas las universidades, si bien en el caso de no poder abarcar todas se podría limitar la investigación a algunas representativas del conjunto. Temporal: Convendría eligir un extenso período que comprendiera distintos períodos de la historia de las universidades españolas, por ejemplo, desde 1890 o desde 1920 hasta la fecha.

Definición del universo: Viene definido por el fin y contenido de la investigación: las actas de las sesiones de apertura del curso universitario. La unidad de análisis serían los documentos de que se trata completos. Se excluirían los discursos meramente técnicos, salvo si tuvieran alguna parte ideológica o moral.

5. Procedimiento de recogida de datos: Se realizaría mediante el examen individual de los documentos base del estudio, tomados de los archivos o bibliotecas de las universidades. En cada documento examinado se determinaría la frecuencia y extensión de las alusiones a cada tipo de valores. Como instrumento de recogida de datos se podría utilizar una ficha para cada documento en el que en la cabecera constarían sus datos generales: Universidad, fecha, clase de documentos, y en el cuerpo se asignarían las filas a las distintas categorías de valores establecidos, y las columnas al espacio y frecuencia de las citas o referencias a cada tipo de valor y a las observaciones pertinentes en cada caso.

(Ejercicio inspirado en el estudio de Vivianne Isambert-Jamati: *Permanence ou variation des objectifs poursuivies par les lycées depuis cent ans.* "Rev. Franç. de Soc.", 1967, número especial.)

89

En las sociedades modernas, y concretamente en España, se ha superado por lo general la antigua costumbre social de la imposición por los padres a sus hijos del cónyuge a elegir. Tampoco está prescrita la homogamia o elección del cónyuge dentro del mismo grupo social. Sin embargo, investigaciones sociales realizadas en diversos países muestran que tampoco existe en la actualidad una libertad completa

de elección de la pareja, sino que ésta viene condicionada por la similitud de ocupación, clase social y lugar de residencia.

Supuesto que se piensa investigar la relación de los matrimonios contraídos en Madrid con la proximidad de residencia y la igualdad de ocupación de los cónyuges, se pide preparar un proyecto esquemático de investigación basado en los documentos estadísticos existentes sobre matrimonios contraídos, que comprenda:

1. La formulación de hipótesis.

2. Su análisis.

3. La operacionalización de las variables a emplear.

4. Delimitación del campo de investigación.

5. Determinación del procedimiento a seguir para la recogida de datos y del instrumento a emplear al efecto.

RESPUESTA

1. Teniendo en cuenta los supuestos teóricos indicados, se podría formular la hipótesis provisional siguiente: Si en una comunidad se toman en conjunto el número de matrimonios contraídos, es probable se encuentre una proporción significativa de ellos en los que se da la similitud de ocupación de los contrayentes o la proximidad de residencia de ambos.

2. Análisis de la hipótesis: Unidad de observación: Contrayentes del matrimonio en una comunidad, en este caso Madrid. Variables. Dependiente: Número de matrimonios contraídos. Independientes: Similitud de ocupación de los contrayentes y proximidad del lugar de su residencia.

3. Operacionalización de las variables. Respecto a la variable ocupación se podría adoptar una división de ocupaciones que se estime adecuada a la finalidad del trabajo. A fin de graduar la similitud de ocupación, se debería ordenarlas, mejor después del trabajo de recogida de datos, según el nivel social que se estime representen.

En cuanto a la variable proximidad del lugar de vivienda de los contrayentes, bastaría con establecer varias categorías, según la mayor o menor distancia. Por ejemplo: 1. Habitación en el mismo domicilio. 2. Vivienda a menos de un kilómetro de distancia. 3. Vivienda de uno hasta tres kilómetros de distancia. 4. Vivienda de tres hasta cinco kilómetros. 5. Vivienda de cinco hasta diez kilómetros. 6. Vivienda a diez kilómetros o más. 7. Vivienda en otra localidad.

4. Delimitación del campo de investigación. Temporal: Dado que suele haber variaciones significativas según los meses en el número de casamientos, el período a investigar podría ser un año como mínimo. Geográfica: A este respecto sería necesario fijar con precisión y concretar en un plano el área de Madrid que abarcaría la investigación. Por el objeto: Todos los matrimonios celebrados. Para facilitar la investigación se podría limitar a los matrimonios canónicos, dejando a un lado los civiles, dado el porcentaje mínimo de éstos en la realidad actual y que no se trata de una circunstancia significativa en principio en relación a los fines de la investigación. Sólo se incluirían los matrimonios en los que uno de los contrayentes por lo menos habite en Madrid. Por tanto, se excluirían los celebrados en la capital, cuando los dos contrayentes vivieran fuera de Madrid.

5. Procedimiento de recogida de datos: Se utilizarían las informaciones obrantes en los Registros Civiles complementadas con las de las parroquias donde tuvo lugar el matrimonio. Para recoger estas informaciones se podría utilizar una ficha destinada a cada matrimonio, dividida en dos partes, una para los datos del novio y otra para los de la novia en la que constasen los datos personales de cada uno de los cónyuges, y estarían consignadas las categorías de ocupaciones adoptadas, así como las referentes al domicilio de su vivienda antes de casarse para marcar con una señal las correspondientes. Sería interesante que, como se ha indicado, las ocupaciones figuren ordenadas jerárquicamente, lo que serviría para señalar en la ficha misma los grados de distancia del nivel ocupacional de los esposos.

17. CUESTIONARIOS

NOCION Y NOTAS DE LA ENCUESTA

La observación por encuesta, entendida esta palabra en sentido restringido y no como averiguación o pesquisa en general, acepción que abarca todos los tipos de observación, consiste en la obtención de datos de interés sociológico mediante la interrogación a los miembros de la sociedad.

Frente a los demás procedimientos de observación, los rasgos que caracterizan a la encuesta son los siguientes:

1. Consistir en la observación no directa de los hechos, como sucede en la observación simple, sino a través de las manifestaciones realizadas por los propios interesados.

2. Ser un método de obtención de datos preparado especialmente para la investigación sociológica, a diferencia de la observación documental, en la que se utilizan con fines sociológicos datos y hechos recogidos de antemano y con otros fines.

3. Permitir, frente al carácter particular, propio de la observación simple y experimental, mucho más limitado en los hechos que puede abarcar, una aplicación masiva que mediante los sistemas de muestreo puede extenderse a comunidades nacionales e internacionales enteras, facultando además para la obtención de informaciones sobre un amplio abanico de cuestiones a la vez y no sólo sobre un aspecto o problema definido.

4. Hacer posible que la investigación social se extienda a los aspectos subjetivos de los miembros de la sociedad y, por tanto, de los hechos y fenómenos sociales.

Todas estas notas, especialmente la amplitud de sujetos y aspectos que puede abarcar, así como la posibilidad de diseñarlo y modelarlo de la manera que se crea más adecuada a los fines pretendidos, hacen de la observación por encuesta el procedimiento sociológico de investigación fundamental y el más empleado de hecho en la realidad.

EL CUESTIONARIO Y SUS CLASES

El instrumento básico de la observación por encuesta es el cuestionario. Este no es otra cosa que un conjunto de preguntas, preparado cuidadosamente, sobre los hechos y aspectos que interesan en una investigación sociológica para su contestación por la población o su muestra a que se extiende el estudio emprendido.

En el cuestionario se pueden distinguir tres clases: el cuestionario simple, la entrevista y las escalas sociométricas. El primero es aquel que los encuestados, previa su lectura, contestan por escrito, sin intervención directa de persona alguna de las que colaboran en la investigación. En las entrevistas el cuestionario es aplicado a los sujetos investigados por personas especializadas en esta tarea, reclutados y preparados por la dirección de la investigación, quienes hacen a los encuestados las preguntas del cuestionario y anotan en él sus respuestas.

Las escalas sociométricas son una forma especial de cuestionario, caracterizada porque las preguntas o las diferentes contestaciones a las preguntas tienen atribuido un valor numérico, lo que permite cifrar cuantitativamente y en cierto modo medir el nivel que alcanza en cada caso la actitud o aspecto investigado.

Hay que advertir que, aunque el cuestionario sea la forma más corriente de las entrevistas y escalas, las primeras se pueden verificar sin sujeción a un cuestionario predeterminado y las segundas también pueden ser independientes de un cuestionario y estar ligadas, como en las escalas de Chapin, Sevell y otros para medir el status socioeconómico, a la observación simple.

El cuestionario y la entrevista representan distintas formas de aplicación del cuestionario. Existe otra forma de administración del mismo, intermedia respecto a las dos anteriores. Consiste en la contestación individual por los encuestados reunidos en un mismo lugar, por escrito bajo la supervisión del investigador.

El cuestionario cumple una función de enlace entre los objetivos de la investigación y la realidad de la población observada. Por ello, las condiciones fundamentales que debe reunir, dependen de la investigación y de la población. Se pueden sintetizar, por una parte en traducir los objetivos de la investigación en preguntas concretas sobre dicha realidad, y por otra parte, en ser capaz de suscitar en los encuestados respuestas sinceras y claras a cada pregunta, que puedan después ser tratadas científicamente, es decir, clasificadas y analizadas.

Así como la observación ocupa un puesto central en el proceso de investigación, igualmente el cuestionario, instrumento en la investi-

gación por encuesta, tiene una importancia central en ella. Así se ha podido afirmar por algunos que una encuesta no puede ser mejor que su cuestionario, ya que, en general, las respuestas no pueden ser mejor que las preguntas, de modo que si estas son vagas, las respuestas serán imprecisas, y si son capciosas, las respuestas serán sesgadas.

La finalidad del cuestionario es obtener de manera sistemática y ordenada, información de la población investigada sobre las variables objeto de la investigación. Esta información generalmente se refiere a lo que las personas encuestadas son, hacen, opinan, piensan, sienten, esperan, quieren o desprecian, aprueban o desaprueban, o a los motivos de sus actos, opiniones y actitudes.

Sistematizando la enumeración anterior, se puede decir con C. Javeau (1971, p. 1), que los objetos ordinarios de las encuestas sociológicas se refieren a tres grandes categorías de datos: 1ª. Hechos (datos actuales) relativos: a) al dominio personal de los individuos que forman el grupo social estudiado: por ejemplo, edad, grado de instrucción. b) al dominio del ambiente que les rodea: por ejemplo, vivienda, relaciones familiares, de vecindad, de trabajo, etc.; c) al dominio de su comportamiento. No podrá tratarse aquí más que, evidentemente, del comportamiento reconocido o aparente. 2ª. Opiniones, a las cuales se juntan los niveles de información, de espera, etc., todo lo que uno podría llamar datos subjetivos. 3ª. Actitudes y motivaciones y sentimientos, todo lo que empuja a la acción, al comportamiento, y está a la base de las opiniones. 4ª Cogniciones, es decir, índices de nivel de conocimiento de los diversos temas estudiados en la encuesta. Revela el grado de confianza a conceder a las opiniones sobre juicios subjetivos.

TIPOS DE PREGUNTAS

El elemento básico del cuestionario, como se deduce claramente de su nombre y definición, son las preguntas. Es por ello, que la bondad de un cuestionario no depende de otra cosa sino de la clase de preguntas empleadas en él y de su adecuada formulación.

Reviste, pues, especial interés en relación al cuestionario analizar los tipos de preguntas existentes y las reglas para su formulación.

Desde el punto de vista de la investigación social, las preguntas de un cuestionario son la expresión en forma interrogativa de las variables empíricas o indicadores respecto a los cuales interesa obtener información mediante la encuesta.

Del mismo modo que las variables tienen que comprender, necesariamente para ser tales, diversos elementos de variación o categorías, igualmente las preguntas de un cuestionario comprenden diversas respuestas,

por lo menos dos, que son los elementos de variación o las categorías de la variable verbalizada por la pregunta.

Las preguntas del cuestionario pues, se subdividen en respuestas, que son los elementos de variación o categorías de la variable a que se refiere la pregunta. Entre ellas el encuestado debe elegir o indicar la que a él le corresponde, la que se ajuste a su caso.

Es fundamental en el cuestionario que las respuestas o categorías que se ofrezcan con las preguntas reunan dos condiciones esenciales: las de ser exhaustivas y excluyentes. (Ver ejercicio 97)

Serán exhaustivas si las categorías o respuestas de las preguntas abarcan todos los casos que pueden darse, de modo que ningún encuestado pueda dejar de responder por no encontrar su categoría.

Serán excluyentes, cuando no pueda darse el caso que un encuestado pueda elegir válidamente dos respuestas distintas de la misma pregunta.

Los tipos de preguntas son múltiples y diversas sus formas de clasificación. Entre otras formas se puede agrupar las preguntas según la contestación que admitan del encuestado, según la naturaleza del contenido de las preguntas, según su función en el cuestionario y según su finalidad.

Según la contestación que admiten del encuestado se distinguen las preguntas cerradas, categorizadas y abiertas.

Las preguntas cerradas son las que solo dan opción a dos respuestas, la afirmativa o la negativa, generalmente sí y no, y, en su caso, no sé y sin opinión.

Las categorizadas, también llamadas preguntas cafetería, presentan como respuestas una serie de categorías entre las que el encuestado debe elegir.

Las preguntas abiertas solo contienen la pregunta y no establecen previamente ningún tipo de respuesta, dejando ésta, por tanto, al libre arbitrio del encuestado.

En general se puede decir que son recomendables las preguntas categorizadas, sobre todo si se consigna en ellas una categoría genérica, por ejemplo, otros ¿cuáles?. Suministran más información que el simple sí y no de las cerradas y evitan el coste y trabajo que supone la categorización posterior de las abiertas.

En concreto, las preguntas cerradas son apropiadas cuando se trata de preguntas muy precisas sobre cuestiones de hecho, o cuando sólo interese a los fines de la investigación conocer el sí o el no, sin más especificación.

Como señalan Schuman y Presser (1981) la controversia entre las preguntas abiertas y cerradas o categorizadas se ha resuelto en la práctica con la victoria de estas últimas, a causa de su eficacia a efectos de la recogida de datos, la clasificación y el análisis.

Sin embargo, dichos autores estiman que las preguntas abiertas son esenciales para conocer el marco de referencia del encuestado y para redactar después las alternativas a ofrecer en las preguntas categorizadas. Por ello, su empleo se estima oportuno sobre todo en el caso de estudios exploratorios o de preencuestas, y cuando no se puede presumir con antelación las posibles opiniones y reacciones de la población a encuestar, ni se conoce bien su vocabulario.

Respecto a las preguntas cerradas, se plantea la cuestión de su formulación equilibrada o no. La formulación se considera no equilibrada cuando la pregunta se refiere de modo expreso sólo a una de las dos alternativas. Por ejemplo:

¿Está Vd. a favor del aborto?

Si – No.

Por el contrario, las equilibradas recogen de modo expreso las dos alternativas. En el caso anterior:

¿Está Vd. a favor o en contra del aborto?

– A favor
– En contra

Este tipo de equilibrio se llama formal para distinguirlo de aquél en el que la alternativa ofrecida implica un contraargumento.

Por ejemplo:

¿Piensa que se debería poder abortar si lo desea una mujer embarazada o estima que no se le debería permitir poner fin a la vida de un niño aún no nacido?

– Debería poder
– No se debería permitir

Los experimentos realizados, aunque no concluyentes, revelan diferencias a veces de quince puntos entre estas distintas formulaciones.

En síntesis, los autores en cuestión estiman que se debe preferir la forma equilibrada a la no equilibrada, pero teniendo en cuenta que si

se ofrecen contraargumentos, éstos pueden contribuir a definir las preguntas y afectar de este modo a las respuestas.

Por último, se ha de señalar que las preguntas categorizadas con muchas alternativas pueden hacer muy difícil el juicio que supone la elección entre ellas, por lo que se recomienda dividirlas en dos preguntas distintas.

En relación a la naturaleza del contenido de las preguntas se tienen las de identificación, de hecho, de acción, de información, de intención, de aspiraciones, de opinión, de expectativas ante el futuro, y de motivaciones y creencias y actitudes. (Ver ejercicio 92).

Las preguntas de identificación, son las que se refieren a las características básicas de las unidades de observación. Por ejemplo, en el caso de las personas, edad, sexo, estado civil, región de nacimiento y de residencia, profesión, estudios, ingresos, religión, filiación política, número de hijos, nacionalidad, raza, etc.

Estas preguntas reciben en inglés el nombre de background, fundamentales, porque, se refieren a las variables independientes principales, respecto de las cuales se puede hacer la hipótesis o suposición que influyen en todas las demás, y en especial en la variable o variables dependientes a estudiar. (Ver ejercicios 94, 120).

Según su función en el cuestionario, se puede hablar, entre otras, de preguntas sustantivas, filtro, de control, de consistencia, de introducción o rompehielos, de alivio, amortiguadores y las baterías de preguntas.

Son sustantivas las preguntas básicas y las referentes a las cuestiones investigadas en la encuesta.

Las preguntas filtro son las que se realizan previamente a otra pregunta, a fin de eliminar a los que no les afecte ésta. Por ejemplo, si preguntamos sin más ¿piensa Vd. comprarse TV. en color? muchos o algunos que contesten que no, puede ser porque ya la tengan. Por lo tanto, antes se debe hacer una pregunta filtro del tenor siguiente: ¿Tiene Vd. TV en color? Si – No. (Ver ejercicio 100)

Las preguntas de control tienen por finalidad asegurarse del interés y buena fe del encuestado y de la veracidad y fiabilidad de sus respuestas al cuestionario. Incluyen respuestas con trampa o falsas, para ver si el encuestado cae en ellas.

Similares a las preguntas de control son las que tienen por objeto comprobar la consistencia de las respuestas del entrevistado. Se trata de preguntas similares, pero redactadas de distinta forma, que se sitúan espaciadas entre sí para ver si sus respuestas a ambas preguntas son congruentes. (ver ejercicios 98, 99).

Preguntas de introducción o rompehielos son las que se hacen para iniciar el cuestionario, e interesar al encuestado en él, o para pasar de un tema a otro, y, en la entrevista, para ganarse su confianza y romper el hielo del desconocimiento.

Preguntas muelle, colchón o amortiguadoras, son preguntas que abordan temas difíciles o escabrosos, formuladas de tal forma que reduzcan su brusquedad y rudeza. (Ver ejercicio 115)

Las preguntas batería son un conjunto de preguntas sobre la misma cuestión, que se complementan y completan unas a otras, enfocando diversos aspectos de ella.

A veces este conjunto de preguntas se disponen en forma de lo que se llama embudo de preguntas, porque se empiezan por los aspectos más generales hasta llegar a los más concretos y precisos. (Ver ejercicio 95)

Según su finalidad se puede hablar de las preguntas directas y las indirectas. Las primeras son las que no buscan descubrir otra cosa que lo que expresan. Las segundas, al contrario, pretenden averiguar algo distinto de los que se desprende de las palabras usadas. (Ver ejercicios 167

Por último, sabe recordar que se pueden utilizar en las preguntas de los cuestionarios, técnicas proyectivas, cuando se quiere lograr una mayor profundidad.

Este tipo de preguntas es útil para captar las ideas, sentimientos, opiniones y actitudes más profundas, a veces, incluso inconscientes para el individuo, que normalmente no se pueden detectar mediante otro tipo de preguntas.

Estas técnicas pertenecen al área de la investigación psicológica-social y como tales no parecen muy adecuadas para su aplicación en las encuestas sociológicas masivas, pues exigen un análisis individual de cada caso pormenorizado, poco compatible con las técnicas de proceso de datos.

No obstante, no se ve imposibilidad de que alguna o algunas de las preguntas del cuestionario, en combinación con otras convencionales, pueden ser aplicación de estas técnicas proyectivas.

Entre estas técnicas se encuentran:

1. La asociación. Se presenta una palabra, frase o dibujo y se pregunta que se exprese lo primero que sugiera tal palabra, dibujo o pregunta.

2. Fantasía. Se invita al encuestado a inventar o contar una historia, o a discutir un dibujo en términos imaginativos.

3. Conceptualización. Se puede descubrir los sentimientos y actitudes del encuestado por la manera en que nombra, ordena y clasifica las cosas y personas.

4. Sentencia incompleta, que el encuestado debe completar con lo primero que se le ocurra. Ejemplo: Mi cabeza ....; si mi padre ....; o simplemente si ...

5. Interpretación libre de dibujos y caritaturas.

6. Presentación de historietas, para que el encuestado elija entre los tipos o conductas que se contienen en ellas.

7. Preguntas pseudoactuales. Son preguntas sobre realidades sociales conocidas. Por ejemplo: ¿Qué tanto por ciento de nazis hay aquí?

8. Adivinar quién son. Son preguntas de este tipo: Existen muchos ladrones en nuestra sociedad. ¿Quienes son?

En fin, también se pueden llamar proyectivas las *chek-list, rating,* inventarios y rejillas a que se refiere Oppenheim (1966, 80 y ss.) Ver ejercicio 120.

REGLAS PARA LA FORMULACION DE PREGUNTAS

En general una pregunta bien formulada, como señala Magdalena Grawitz (1975, p. 246), es aquella que a) no ejerce influencia en el sentido de la respuesta. b) no incita a una respuesta inexacta, que no corresponde a la información buscada.

La influencia ejercida por la pregunta puede provocar, desde el error, si la pregunta redactada en términos ambiguos es mal interpretada, hasta la orientación más o menos fuerte del encuestado, en un sentido determinado. Esta infuencia recibe el nombre de sesgo, *bias* en la terminología anglosajona.

Existen algunas reglas, fundadas en el sentido común y la experiencia simple, que contienen criterios a seguir en la formulación de preguntas de un cuestionario.

Como todas las reglas de este tipo no tienen un valor absoluto, sino indicativo. En último término, es el equipo investigador el que debe decidir en cada caso las preguntas a formular según las circunstancias de la investigación, entre las que cabe destacar su objetivo, la información que interese recoger y el público que se va a encuestar.

Entre las reglas para la formulación de las preguntas del cuestionario on muy conocidas las siguientes, formuladas por Bowley:

1. Las preguntas han de ser relativamente pocas.

2. Las preguntas han de estar hechas de tal forma que requieran siempre una respuesta numérica o simplemente una afirmación o negación, o bien la elección de una categoría propuesta.

3. Las preguntas han de ser sencillas y redactadas de tal forma que puedan comprenderse con facilidad por las personas a las que van destinadas.

4. Las preguntas han de estar hechas de forma que no levanten prejuicios. (Ver ejercicio 106).

5. Las preguntas no deben ser indiscretas sin necesidad.

6. En lo posible, las preguntas que se hagan han de ser corroborativas. (Ver ejercicio 137).

7. Las preguntas habrán de estar hechas en forma que contesten directa o inequívocamente al punto de información deseado.

Otros criterios a tener en cuenta a este respecto se pueden señalar:

No hacer directamente preguntas que obliguen a cálculos o esfuerzos de memoria del encuestado, antes bien, en este caso se deben presentar de modo que se facilite la memorización. (Ver ejercicios 112 a 114).

Redactar las preguntas en forma personal y directa, no impersonal. No realizar preguntas embarazosas que puedan dejar al descubierto o poner en evidencia al encuestado y si hay necesidad de hacerlas presentarlas de forma que no hieran su sensibilidad. (Ver ejercicios 115, 116).

No presentar las preguntas en forma negativa, pues dan lugar a duda sobre el sentido de la respuesta, ni tampoco en forma afirmativa, pues sugieren el contestar afirmativamente. Es conveniente formularlas en forma neutral, ni afirmativa ni negativa. (Ver ejercicios 105, 107).

No utilizar sin precisión palabras abstractas, por ejemplo, clase, especie, tipo, etc., ni tampoco de tipo valorativo: mucho, escaso, bueno, malo, etc., pues se pueden tomar en diferentes sentidos y cada uno las concede diverso alcance. (Ver ejercicio 103).

Formular las preguntas del modo más concreto y preciso posible, sobre todo en el caso de que se refiera a unidades de medida o períodos de tiempo. (Ver ejercicio 102).

Evitar el empleo de la interrogación por qué. (Ver ejercicio 107).

Procurar que las preguntas, si son abiertas no presenten una opción alternativa, ni comprendan en realidad dos preguntas y que ni siquiera contengan aclaraciones de su sentido, porque en estos casos se puede dar lugar a confusión en el encuestado y se oscurece el sentido de las respuestas. (Ver ejercicios 104, 110 y 118)

Es muy importante que las preguntas sean cortas en lo posible, para evitar que sean ambiguas y oscuras.

En fin, tener en cuenta al formular las preguntas el marco de referencia del encuestado, es decir, la situación social en que se halla que ha de constituir el punto de vista de su respuesta.

Por último, tan meticuloso debe ser la preparación del cuestionario que no se debe olvidar que las palabras que se empleen en las preguntas han de ser elegidas cuidadosamente.

Payne *(The art of asquing quaestions. Princeton Univ. Press)* aconseja preguntarse en relación con las palabras:

1. Si significan exactamente lo que se quiere decir.

2. Si no tienen otro sentido, y en este caso, si el contexto lo precisa suficientemente.

3. Si no se prestan a confusión (oralmente o por escrito) con otra palabra.

4. Si no puede encontrarse otra palabra u otro giro más sencillo y claro.

PREPARACION DE CUESTIONARIO. ETAPAS

La construcción de un cuestionario es una operación muy delicada y difícil. Es delicada, ya que, dada la importancia central del cuestionario en la encuesta, importa que sea realizada con la mayor perfección porque de ello depende el éxito de la encuesta.

Es difícil porque además de la gran dificultad que entraña siempre en la investigación social el captar y expresar en indicadores concretos los aspectos de la realidad, significativos en relación al objeto de la investigación, implica también muchos problemas de otro tipo, entre los que cabe destacar los lingüísticos y psicológicos.

Este carácter delicado y difícil de la construcción de un cuestionario, exige que normalmente su preparación deba ser una labor de equipo.

La investigación por encuesta debe traducir las variables empíricas sobre las que desee obtener información en preguntas concretas sobre la realidad social a investigar, capaces de suscitar respuestas sinceras y claras.

Supuesto lo anterior, las etapas que se pueden distinguir en la preparación del cuestionario son las siguientes:

1. Formulación de hipótesis.

2. Determinación de las variables a observar con sus dimensiones e indicadores.

3. Planificación del contenido del cuestionario.

Como etapa previa a éstas, hay que mencionar la preencuesta que consiste específicamente, en la realización de un primer contacto, por entrevistas no directivas y de grupo, con la población que se va a estudiar para conocer su situación, mentalidad, problemas, etc.

FORMULACION DE HIPOTESIS

En esta etapa de formulación de hipótesis se pueden distinguir tres niveles:

El primero, el establecimiento de la hipótesis o hipótesis generales elegida o elegidas como respuestas probables a la cuestión general investigada.

El segundo, la especificación de las hipótesis generales en subhipótesis que las desarrollen y expresen en los distintos aspectos que comprendan.

El tercer nivel es el más propio y específico del cuestionario. Consiste en la concepción de lo que se va a buscar con cada pregunta, de lo que se pretende saber o conocer por su mediación, o sea prever su sentido y utilidad.

Que las preguntas del cuestionario impliquen hipótesis en este sentido es la condición para que estas tengan un sentido preciso, respondan a una necesidad en relación a los objetivos de la encuesta y sean, en último término, útiles y beneficiosas para los fines perseguidos en el cuestionario.

Incluso las categorías de cada pregunta deben ser establecidas inteligentemente. Esto quiere decir que deben obedecer a una intención determinada, o lo que es lo mismo, que deben fundarse también en hipótesis. (Ver ejercicios 90, 130).

DETERMINACION DE LAS VARIABLES

Todo trabajo de investigación que se nos propone es en su tema o título una variable general. A fin de que pueda ser observada en la rea-

lidad es preciso operativizarla o hacerla operativa, especificando sus dimensiones e indicadores.

Igualmente en la investigación social por cuestionario, el tema sobre el que versa se refiere también a una variable general. Es preciso, pues, como se acaba de indicar, hacerla operativa determinando sus variables intermedias y los indicadores.

La operativización de variables, pues, puede y debe servir también para determinar los puntos de información a obtener, mediante la encuesta y que han de ser desarrollados dentro del cuestionario mediante las correspondientes preguntas.

Como es claro estos puntos de información que resulten de la operativización se deben someter a juicio crítico sobre la oportunidad de su inclusión en el cuestionario. Además se deben completar con aquellos aspectos no derivados de esta operación y que se crean convenientes y de interés.

Todos estos puntos de información pueden constituir la base de lo que se llama guía del cuestionario, que se suele utilizar como orientación para la redacción del mismo, y en la que es lógico se comprendan también las decisiones a que lleguemos sobre la planificación del cuestionario.

PLANIFICACION DEL CONTENIDO DEL CUESTIONARIO

La finalidad de ésta etapa es:

1. La especificación del tipo de preguntas más adecuadas, cerradas, categorizadas, etc., para obtener de la población encuestada los datos o informaciones que señalan los indicadores establecidos.

2. Las categorías, en su caso, que más convengan a cada pregunta y a los objetivos que se persiguen con ellas.

3. La determinación del número de preguntas y de su orden y disposición en el cuestionario.

Los dos primeros pasos se deben realizar teniendo en cuenta:

a) Las hipótesis concretas previstas, es decir, lo que se busca con cada pregunta y sus categorías, o sea, lo que se pretende saber o conocer con ella.

b) Las características de la población estudiada.

c) La naturaleza y aplicaciones de cada tipo de pregunta.

d) La forma de aplicación del cuestionario, por escrito o mediante entrevista.

El destino de los datos derivados de cada una de las categorías de las preguntas es su agrupación y sistematización en tablas.

En una tabla, a cada respuesta debe corresponder una columna. Por ello, hay que tener en cuenta en la planificación del cuestionario que un número excesivo de respuestas da lugar a tablas de muchas columnas y por tanto difíciles de leer y analizar.

En cuanto al tercer paso, número y orden de las preguntas se pueden recomendar las siguientes reglas:

1. El cuestionario debe contener todas las preguntas necesarias, pero ninguna más. Un cuestionario largo, salvo caso de absoluta necesidad, es penoso tanto para el entrevistador como para el encuestado.

2. No se debe empezar el cuestionario con preguntas difíciles y embarazosas.

3. Se debe evitar a toda costa colocar juntas preguntas en las que la respuesta a una de ellas pueda influir en el sentido de la respuesta a la otra.

Si esto ocurre, se produce el efecto de "halo" o "contagio". (Ver ejercicios 101, 107).

4. Se debe procurar que todas las preguntas sobre un aspecto figuren juntas en el cuestionario, dispuestas así como los mismos temas, guardando un orden temporal, lógico y psicológico.

5. Para evitar la monotonía y en cansancio y facilitar los cambios de tema, es conveniente intercalar en sitios estratégicos preguntas de alivio más o menos intrascendentes, que supongan un cierto respiro.

6. Se ha observado, que, en el caso de las entrevistas, la curva de tensión de los entrevistados se manifiesta plana al principio y al final, y suele alcanzar su máxima intensidad en el segundo tercio de la entrevista.

Por esta razón, las instituciones de investigaciones de mercado suelen comenzar las entrevistas con preguntas interesantes para el entrevistado, que exigen una motivación menor y terminan con las preguntas de identificación que requieren menos esfuerzo.

La tendencia es, pues, a incluir las preguntas más importantes en el tercio medio de la entrevista, y a no hacer preguntas que puedan comprometer al encuestado, sino después que se haya establecido una cierta relación de confianza con él. (Ver Ejercicio 130).

Se considera que en temas importantes para la investigación es peligroso basarse en una sola pregunta. Varias preguntas de diferente forma, redacción y contenido reducirán el peligro de los diversos sesgos a que está sometida una sola pregunta.

Similarmente, se estima que las preguntas referentes a actitudes, o posiciones permanentes y profundas de la persona, deben ser acompañadas por lo menos de otra pregunta que mida la fuerza de la actitud en comparación con otros sentimientos o actitudes, es decir, su intensidad, o bien su centralidad, o sea, la importancia que se concede a la actitud expresada a la hora de tomar una decisión, por ejemplo, votar.

ELABORACION FORMAL DEL CUESTIONARIO

Planificada del modo indicado la construcción del cuestionario, es preciso proceder a su elaboración formal. Destacan en ella la redacción formal de las preguntas y la determinación de los aspectos formales del cuestionario y de sus elementos accesorios.

La redacción de las preguntas, según la tesis central del libro de Sudman y Badburn (1982) es el elemento esencial para maximizar la validez de los datos obtenidos en la investigación por encuesta. Ya desde los primeros tiempos de la encuesta, se conoce el hecho de que cambios aparentemente pequeños en la redacción de las preguntas pueden causar grandes diferencias en los diversos porcentajes de respuestas a la misma.

Reglas básicas en esta redacción son que no se pierda nunca de vista los objetivos de la investigación y que, previamente a la formulación de cada pregunta, nos preguntemos sobre la necesidad y oportunidad de la misma, y sobre el grado en que nos servirá para lograr los objetivos propuestos.

Finalidad de los aspectos formales es que el cuestionario sea atractivo y adecuado a las características de la encuesta. Su objeto es la programación de sus condiciones formales, tales como papel a emplear, clase, color, tipo de impresión; espacios destinados a las respuestas y a la codificación; márgenes a dejar a derecha e izquierda y en medio de las preguntas y datos de identificación que deba contener el cuestionario.

Como elementos accesorios al cuestionario cabe destacar:

a) La carta dirigida al encuestado solicitando su cooperación a la encuesta rellenando el cuestionario y dándole las gracias por su cooperación.

b) Las instrucciones para cumplimentar el cuestionario.

La carta deberá estar escrita de forma clara y sugestiva, y en lenguaje sencillo y apropiado al destinatario. En ella se hará resaltar el interés de la investigación y los beneficios directos e indirectos que representa su contestación.

Las instrucciones, que según Zarkowich (1970), son tan importantes como el cuestionario, son el medio de obtener la normalización de las respuestas recogidas y de evitar la introducción de elementos subjetivos en las mismas.

En ellas se expondrán las advertencias, que se consideran necesario o conveniente realizar, sobre la forma general de cumplimentar el cuestionario y de consignar las respuestas según los distintos tipos de preguntas que comprenda.

Las instrucciones se deben redactar con gran claridad y esmero. Se debe procurar que sean completas y resuelvan todas las dudas que puedan plantearse. Sobre todo ha de tenerse cuidado especial en no inducir a confusión. En este caso podrían ser causa de sesgo.

JUICIO CRITICO DE UN CUESTIONARIO

Este juicio se puede dividir en dos partes, una sintética y otra específica.

Respecto a ambas se suelen destacar los siguientes puntos:

Sintética. En ella se examinará:

1. La adecuación de las preguntas o items del cuestionario a la finalidad u objetivo de la investigación.

2. La estructura y disposición equilibrada y armónica o no de sus partes.

3. Se verá si no falta ninguna parte o conjunto de preguntas esenciales al fin que se pretende.

4. Se examinará si es más largo de lo debido y si algunas preguntas son reiterativas o superfluas.

Específica. Se analizará cada ítem por separado indicando el carácter de la pregunta, su tipo, su función en el cuestionario y si su formulación se ajusta a las reglas enumeradas. (Ver ejercicios 119, 130).

LOS CUESTIONARIOS NO CUMPLIMENTADOS

Tanto en el caso del cuestionario simple como en el de la entrevista, hay por lo general siempre un número más o menos elevado de cuestionarios que no son devueltos, o de entrevistas que no se pueden efectuar.

En los cuestionarios por entrevista es frecuente el porcentaje de las que no se pueden efectuar, oscilen entre el 5 y el 15 % y que se eleven al 40 % o más en los cuestionarios enviados por correo.

El hecho de que un porcentaje relativamente elevado de cuestionarios se frustre, plantea una problema al equipo investigador, porque se puede suponer que los que no contestan, precisamente por ello, son individuos de la población de características diferentes a los que contestan. Su no contestación implica que una parte de la población, probablemente distinta de los que contestan, no está representada en la muestra, lo que introduce un sesgo o fuente de error en la misma.

Este sesgo de la muestra no se puede remediar aumentando su tamaño, para sustituir a los que no contestaron. Esta solución no elimina el hecho probable aludido de que los que no contestan son probablemente gente distinta de los que contestan, que seguramente de haber contestado lo habrían hecho de modo distinto.

El único procedimiento que se considera válido para solucionar este problema es la repetición de visitas a los que no quieran contestar a la entrevista y el recordatorio o envío de nuevos cuestionarios por correo a los que no lo han devuelto, hasta conseguir reducir al mínimo los cuestionarios no cumplimentados.

Otro procedimiento también válido es el llamado de Hansen que consiste en formar una submuestra con parte de los que no han contestado el cuestionario, en representación de todos éstos. Luego hacer todo lo posible cueste lo que cueste, hasta conseguir su respuesta. Esta, se considera representativa de todo el grupo y que por tanto se extiende a él. (Ver ejercicio 151).

También es factible la ponderación en los resultados de la encuesta del conjunto de no-respuestas, lo que tiene por objeto lograr que la muestra no pierda representatividad por este motivo, así como la imputación la cual pretende dar valores de respuesta a los encuestados que no contestan a alguna de las preguntas del cuestionario, (Cf. Sánchez Carrión 1995, 207 y ss.).

LA NO RESPUESTA

Otro problema de los cuestionarios está constituido por la falta de respuestas a las preguntas. Esta falta de respuesta o no respuesta puede adoptar dos modalidades:

1. La de los que contestan no sé y
2. La de los que no dan respuesta alguna a la pregunta o no contestan.

Existe tendencia a pasar por alto las no respuestas y a concederlas poca importancia. Sin embargo, sobre todo en el caso de los no sé, implican una modalidad de respuesta, que merece la atención de los investigadores.

El no contesta demuestra una actitud que hace que el encuestado, por neutralidad o indiferencia, o por la razón que sea, se reserve su opinión y se niegue a manifestarla y a colaborar en la encuesta respondiendo a la pregunta.

En cambio, los no sé pueden tener distinta significación según los casos, que depende de los tipos diversos (Zeisel, 1974) de no sé. En primer lugar, se puede distinguir los que Zeisel llama "no sé" auténticos, que indican una falta de conocimiento de lo preguntado y, por tanto, una incapacidad de dar una respuesta a la pregunta. Por ejemplo, quién es el director de tal película? En este caso, la respuesta no sé, forma parte de la información que se desea obtener mediante la pregunta.

El no sé puede significar también tanto que no se está familiarizado con la cuestión hasta el punto de dar una respuesta, como que, aunque se esté, no se tiene una actitud definida respecto al mismo. ¿En su opinión es cierto que el café tal no perturba el sueño?

En segundo lugar, están los no sé que indican un fallo en la respuesta. Es el caso de preguntas sobre la conducta del encuestado, de lo que hace, como, por ejemplo ¿vió Vd. la TV ayer o en la última semana? ¿Fue a misa el último domingo? Este tipo de no sé pueden ser reducidos mediante la actuación del entrevistador.

El no sé, en fin, puede significar una evasiva intencional, cuando se trate de preguntas en la que los encuestados suponen que las respuestas que den pueden ir en detrimento suyo.

Para interpretar los no sé, se puede suponer, cuando falta toda otra información, que se reparten al azar respecto a las demás categorías de la pregunta y, por tanto, se pueden distribuir proporcionalmente entre ellas.

Respecto a la inclusión de la categoría no sé, existen dos posiciones (Schuman y Presser, 1981):

1. La que recomienda evitar dicha inclusión como una categoría expresa de respuesta a las preguntas, y que se consigne únicamente si el entrevistado la formula espontáneamente.

2. La de los que ven en el no sé una categoría de respuesta tan legítima como las demás e incluso sostienen que se debe alentar.

En todo caso, es un hecho comprobado que la inclusión expresa de la categoría no sé aumenta significativamente el % de los que dan esta respuesta a las preguntas.

VALIDEZ Y SEGURIDAD DEL CUESTIONARIO

En el cuestionario, como en todos los instrumentos de observación, además del pretest o prueba a que ya hemos aludido, para comprobar que las preguntas son comprensibles y que las respuestas a ellas son significativas, en orden a la investigación pretendida, es necesario también verificar su seguridad y validez,

Un procedimiento para comprobar la seguridad del cuestionario es el llamado retest, que consiste en aplicaciones repetidas de aquél a las mismas personas, dentro de un breve espacio de tiempo.

En cuanto a la validez, el único procedimiento seguro para verificarla, es comprobar que los resultados obtenidos se ajustan a la realidad. Pero esto en los cuestionarios sólo es posible en las preguntas de hecho, pero no en las de opinión.

Como procedimientos prácticos para comprobar indirectamente la validez de los cuestionarios, se pueden utilizar las técnicas de los grupos conocidos, "know groups" y la de "predictive validity", la "random probe" o prueba aleatoria y las "cross-check-questions".

Según la primera técnica, el cuestionario se aplica a individuos que se sabe de antemano presentan unos valores muy altos o muy bajos respecto a determinadas características, que se pretenden conocer con el cuestionario. Si los resultados de esta aplicación coinciden significativamente con los ya conocidos es un índice de validez.

En la segunda técnica, se observa la conducta posterior de los encuestados y se contrasta con sus opiniones. Si son congruentes ambas, también se puede tomar con índice de validez del cuestionario.

La "random probe" o prueba aleatoria consiste en preguntas situadas aleatoriamente en el cuestionario. Tienen como finalidad precisar y desarrollar lo que las preguntas significan para el encuestado. Sirven para aclarar sus respuestas y para saber el grado de comprensión de la encuesta y la validez de sus respuestas.

Un final contraste de validez es el uso de "cross-check questions". En este caso se incluyen preguntas que pueden ser contrastadas con fuentes

independientes. Por ejemplo, la pregunta ¿cuántos años ha estado en tal prisión?, puede ser contrastada con el registro de la prisión.

Respecto a la validez, en general, de la encuesta es aplicable lo expuesto en el capítulo 7 sobre los tipos de diseño y las variables extrañas las cuales, si se producen, pueden influir en los encuestados y alterar los resultados. De modo específico se debe evitar caer en los defectos de redacción y en el orden de las preguntas a los que se ha hecho mención a lo largo de todo este capítulo, pues todos ellos suponen sesgos que afectan a la validez.

El mejor medio para asegurar la validez de los cuestionarios es procurar reducir al mínimo los errores de todo tipo.

ETICA DE LA INVESTIGACION POR ENCUESTA

Toda investigación es, en sí, por ser tal, una inquisición, que cuando se trata de encuesta ha de versar necesariamente sobre personas, e implica una intromisión en la intimidad de sus condiciones de vida, conducta, opiniones y actitudes. Por ello plantea el problema de los límites de su licitud, es decir, de su ética.

Con relación a esta ética, hay que partir del principio de que el investigador social no tiene derecho a inmiscuirse en la vida de las personas, sin su conocimiento y aceptación totalmente consciente y libre. Tampoco el investigador social tiene derecho a hacer uso de las informaciones que obtenga de la encuesta, salvo estrictamente para los fines conocidos y autorizados por los encuestados.

De estos principios se derivan tres exigencias éticas básicas en la encuesta:

1. Que se informe a los encuestados del origen, razón de ser, entidad que promueve la encuesta, su objeto y finalidad, utilización de los datos y, en general, de todos los aspectos que pueden influir en su participación.
2. Que se respete la plena libertad de los encuestados de participar o no en la encuesta y de abandonarla en cualquier momento.
3. Que se haga con los datos el uso autorizado por los encuestados, y se observe rigurosamente en todo lo demás el secreto de las respuestas y el anonimato de los informantes.

En síntesis, es responsabilidad ética de los investigadores por encuesta (Sudman y Bradburn, 1982, 11) informar a los encuestados de los

objetivos de la encuesta tan completamente como sea apropiado, explicar el contenido general de las preguntas y contestar toda pregunta que el encuestado pueda hacer sobre quién patrocina la encuesta o sobre el uso de los datos obtenidos.

Además, deberá informarle del grado en que se mantendrá la confidencialidad de sus respuestas y se esforzará todo lo posible por asegurarle que este grado de confidencialidad será mantenido.

EJERCICIOS

90

Dadas las siguientes series de preguntas procedentes de los cuestionarios que se indican, se pide determinar las hipótesis que pretendían verificar los promotores de las encuestas con ellas.

A) ¿Cuál es su profesión?

¿En qué trayecto ha utilizado usted las líneas aéreas Air-Inter?

En .....
En ninguno.
No responde.

(En el caso de que los haya utilizado en algún trayecto.)

¿Qué piensa usted de los tipos de aparatos que Air-Inter mantiene en servicio en las líneas que usted ha utilizado?

RESPUESTA

Es admisible afirmar que las hipótesis implícitas en la primera y segunda cuestión son que el viajero tiene alguna profesión y que probablemente ha utilizado antes los aviones Air-Inter en algún otro viaje, si bien se trata de preguntas, la primera, de tipo fundamental, para conocer las condiciones personales del encuestado, y la segunda, de filtro o de identificación, o sea, hecha con la finalidad de determinar si el encuestado pertenece o no al universo de la encuesta.

La hipótesis implícita en la tercera pregunta consiste en que, si el viajero ha notado algún defecto o imperfección en los aparatos en cuestión, lo denunciaría, siendo el conocimiento de estos defectos precisamente el objetivo perseguido con la pregunta.

B) Encuesta por entrevista, previa a unas elecciones legislativas, con el objeto de evaluar las probabilidades de los diputados salientes. La entrevistadora muestra una fotografía que pertenece a un diputado saliente local y pregunta simplemente:

¿Conoce usted a este señor?

Si.
No.

RESPUESTA

La hipótesis implícita es que el reconocimiento visual del diputado saliente es indicio de su popularidad y de la irradiación de su persona y de su acción como tal diputado.

C) Encuesta a los alumnos de escuelas de artes y oficios por cuestionario autoadministrado, con objeto de conocer la "mentalidad social" de esta categoría de jóvenes:

"Si yo comparo mi vida con la de un muchacho que prepara su bachillerato:

Considero que tiene más suerte que yo.

Considero que tiene menos suerte que yo.

Ambas suposiciones son equivalentes.

Sin respuesta."

RESPUESTA

La hipótesis implícita es la de que estos alumnos se comparan con los de bachillerato cuando tratan de evaluar sus posibilidades para el porvenir.

(Basado en R. Mucchielli, 1974, págs. 153-54.)

91

Formular dos preguntas, una abierta y otra cerrada, en el supuesto de un cuestionario aplicado a empleadores y empresarios agrarios sobre si se dedica a otra actividad profesional, y a cuál, teniendo en cuenta que las preguntas cerradas sólo dan opción a dos respuestas contrapuestas, generalmente sí o no, mientras que en las preguntas abiertas queda la contestación al arbitrio del encuestado.

RESPUESTA

Pregunta cerrada:

¿Aparte de trabajar en el campo, se dedica usted a otra cosa? (.)

Si.
No.
No contesta.

Pregunta abierta en el caso de contestar si:

¿Qué tipo de actividad es? (.)

.....

No contesta

(Del primer informe Foessa.)

(.) Se llama la atención sobre las palabras "cosa" y "tipo de". Dado su carácter abstracto y genérico, parece que la primera debería ser sustituida por trabajo y la segunda suprimida.

92

En las siguientes preguntas, tomadas de cuestionarios de encuestas realizadas en España sobre la calidad de vida y la cuestión regional española:

1. *¿A qué horas se siente Vd inseguro?*

   Por la mañana
   Por la tarde hasta las diez.
   A partir de las diez de la noche.
   A cualquier hora.

2. *La ciudad en que habita tiene*

   Más de 500.000 habitantes
   De 100.000 a 500.000
   De 50.000 a 100.000
   De 20.000 a 50.000
   De 5.000 a 20.000
   De 1.000 a 5.000
   Menos de 1.000

3. *Por favor, dígame ¿cónoce Vd. lo que son los fueros?*

   Sí
   No

4. *En caso afirmativo ¿cree que la existencia de los fueros está justificada actualmente?*

   Está justificada
   No lo está
   No sabe

5. *¿Cómo explicaría a un extranjero lo que son los fueros?*

Se pide:

A) Indicar el tipo a que pertenecen por: a) la contestación que admiten del encuestado, b) según la naturaleza de su contenido y c) según su función.

RESPUESTA

1) a) categorizada
   b) de hecho y no de acción, porque es algo que le ocurre a uno y no que él haga.
   c) sustantiva, porque la seguridad ciudadana es uno de los elementos de la calidad de vida.
2) a) categorizada
   b) de identificación
   c) sustantiva;
   Es una de las preguntas básicas del cuestionario.
3) a) cerrada
   b) de conocimiento
   c) de filtro: culple la fuhción, respecto a la pregunta siguiente, de descartar a los que no sepan lo que son los fueros.

4) a) cerrada
   b) de opinión
   c) sustantiva.
5) a) abierta
   b) de conocimiento
   c) de control, sobre si las respuestas 3 y 4 se fundan en un conocimiento exacto de los fueros o no.

B) Señalar el nombre de las variables que verbalizan y el tipo a que pertenecen estas variables por su naturaleza, nivel de abstracción y escala de sus categorías.

RESPUESTA

1. Nombre: Horas en que la gente se siente insegura. Variable: cualitativa y de escala nominal. Por su nivel de abstracción, esta pregunta y las siguientes y, en general, toda pregunta de un cuestionario, es un indicador porque la formulación de la pregunta implica su concreción.

2. Nombre: Habitat. Tipos: cualitativa, de escala nominal porque aquí los números no indican orden, ni la existencia de una unidad de medida del habitat, sino categorías distintas del lugar en que se vive: pueblos pequeños, grandes, etc.

3. Conocimiento de los fueros. Cualitativa y nominal dicotómica.

4. Opinión sobre la justificación de la existencia de los fueros. Cualitativa y nominal dicotómica.

5. Modo de explicación a un extranjero de los fueros. Cualitativa y nominal.

93

Formular una pregunta en forma abierta, cerrada y categorizada para conocer el juicio de los encuestados sobre el Concilio Vaticano II, e indicar su tipo y cuál es la forma más aconsejable en principio.

RESPUESTA

Se trata de una persona de opinión que busca conocer el criterio o juicio de los encuestados sobre acontecimientos, personas, hechos, etc., discutibles de trascendencia social.

Es aconsejable que se presente en principio, de forma categorizada, porque la forma cerrada no matiza nada la opinión, y la abierta puede dar lugar a una gran diversidad de respuestas difíciles de clasificar.

Formulación Cerrada:

¿Qué opinión le merece el Concilio Vaticano II?

Buena.
Mala.
Sin opinión.
No contesta.

¿Puede decirnos su opinión sobre el Concilio Vaticano II?

.....

Sin opinión.
No contesta.

¿Cuál de las respuestas que siguen refleja mejor su opinión respecto al Concilio Vaticano II?

Representa un retroceso espiritual.
Supone el germen de la destrucción de la religión.
No ha solucionado nada.
Se ha contentado con establecer reformas accesorias, sin modificar nada importante.
Implica reformas importantes necesarias en la Iglesia, pero se ha quedado a mitad de camino.
Contiene las bases para una transformación profunda de la Iglesia de acuerdo con las necesidades del mundo moderno.
Otras.
Sin opinión.
No contesta.

94

Se ha proyectado realizar una investigación cuya finalidad es estudiar la distribución de las enfermedades mentales según la clase social.

Se pide formular las preguntas de situación, teniendo en cuenta que, como es sabido, las preguntas de situación, o "background" –fundamentales– en lengua inglesa, son las que sirven para obtener los datos básicos de la persona encuestada. No existe una lista fija de ellas. Su número depende del tipo del estudio y del criterio de los investigadores. En general, se pueden considerar como tales las referentes a la edad, sexo, estado civil, estudios realizados, situación económica, lugar de nacimiento, nacionalidad, raza, profesión, religión y filiación política. Tienen gran importancia en las investigaciones sociales, pues suelen constituir las variables independientes más significativas, a efectos de deducir consecuencias sociológicas en un estudio poniéndolas en relación las demás variables.

Sobre su colocación en el cuestionario, es frecuente que aparezcan de hecho al principio, si bien hay autores que opinan deben colocarse al final, y empezar con preguntas formularias que tiendan a iniciar el diálogo y a lograr la apertura del encuestado.

Es obvio que a veces no será necesario realizar en absoluto alguna de estas preguntas de situación, como cuando la encuesta se dirige sólo a nacionales, o a solteros, o a personas del sexo masculino, etc.

RESPUESTA

Le rogamos tenga a bien contestar a las siguientes preguntas personales de gran interés para este trabajo:

¿Cuál es su sexo?

Varón.
Mujer.

¿Indique por favor en qué grupo de edad de los que siguen está comprendido usted?

Menos de veinte años.
De veinte hasta treinta y cinco.
De treinta y cinco hasta cincuenta.
De cincuenta hasta sesenta y cinco.
De sesenta y cinco hasta setenta y cinco.
De setenta y cinco años o más.

¿Su estado actual es?

Soltero.
Casado, con hijos.
Viudo, con hijos.
Casado, sin hijos.
Viudo, sin hijos.
Otro. ¿Cuál?

¿Cuál es su nacionalidad?

Español.
Extranjero.

¿Cuál es su raza?

Blanca.
Negra.
Amarilla.
Otra.

El lugar donde reside habitualmente tiene una población:

De menos de 1.000 habitantes.
De 1.000 hasta 5.000.
De 5.000 hasta 10.000.
De 10.000 hasta 50.000.
De 50.000 a 100.000.
De 100.000 a 500.000.
De 500.000 o más habitantes.

¿Qué estudios ha realizado?

Ninguno.
Primaria.
Formación profesional artesana o industrial.
Bachillerato.
Universidad primer ciclo.
Licenciatura.
Ingeniería técnica.
Ingeniería superior.

Doctorado.
Otros. ¿Cuáles?

Sus ingresos netos al mes por todos los conceptos son:

Menos de 10.000 pesetas.
De 10.000 hasta 15.000.
De 15.000 hasta 30.000.
De 20.000 hasta 30.000.
De 30.000 hasta 45.000.
De 45.000 hasta 70.000.
De 70.000 hasta 100.000.
De 100.000 pesetas al mes o más.

¿Cuál es su profesión principal?

Obrero del campo.
Trabajador de la industria y servicios.
Cuadros medios y empleados.
Funcionario público.
Militar.
Propietario agrícola.
Profesional liberal.
Director o alto ejecutivo de empresas.
Empresario.
Otra. ¿Cuál?

Desdé el punto de vista religioso, se considera usted:

Católico practicante.
Católico no practicante.
No católico, pero de otra religión.
Escéptico.
No creyente.

¿Qué posición política de las que siguen considera más próxima a la suya?

Extrema izquierda.
Izquierda.
Centro.
Derecha.
Extrema derecha.
Indiferente.

95

Formular una batería de preguntas referentes a la participación en asociaciones voluntarias, recordando para ello que una batería es un conjunto de preguntas que completan, explican y detallan una principal. Por tanto, se debe formular primero la pregunta principal y después las secundarias que las desarrollen.

RESPUESTA

¿Pertenece usted o no a alguna asociación voluntaria?

Sí.
No.

En caso afirmativo, ¿de cuántas forma parte?

De una.
De dos.
De tres.
De cuatro.
De cinco.
De seis.
De seis a nueve.
De nueve a doce.
De más de doce.

Las asociaciones voluntarias a que pertenece son de tipo:

	Número de asociaciones
Deportivo.	
Cultural.	
Religioso.	
Político.	
Recreativo.	
Otras.	

¿Qué grado de participación tiene en ellas?

Clase de asociación	*Asiste raras veces a sus actividades o nunca (número)*	*Asiste casi siempre o siempre (número)*	*Contribuye económicamente (número)*	*Tiene cargos directivos (número)*
Deportivas.				
Culturales.				
Religiosas.				
Políticas.				
Recreativas.				
Otras.				

96

En una encuesta por cuestionario a estudiantes de Sociología se quiere conocer las razones por las que eligieron esta carrera; se pide indicar cómo se deberá proceder en la formulación de las preguntas que se estimen procedentes y formularlas.

RESPUESTA

En los casos que como éste se orientan a investigar los motivos de una determinada elección, conducta, creencia, no se debe limitar uno a preguntar en el cuestiona-

rio, por ejemplo: ¿Por qué eligió usted la carrera de Sociología? Una pregunta similar tendría los inconvenientes de las preguntas tipo ¿por qué?, pero sobre todo no serviría para descubrir los diversos aspectos de la decisión indagada, origen, tendencias personales, influencias y demás circunstancias que intervienen en una acción.

Por ello, la solución más recomendable en este caso sería formular una serie de preguntas orientadas a detectar los aspectos más importantes de la elección adoptada.

¿Cuándo empezó usted a sentir inclinación por la sociología?

.....

¿Consideró usted la posibilidad de elegir otra carrera?

Sí.
No.

En caso afirmativo ¿qué le inclinó hacia la sociología?

¿Qué características de la sociología le atrajeron hacia ella.?

¿Pensó usted que tenía aptitudes para la sociología?

Sí.
No.

¿Quién o quiénes ejercieron influencia en usted sobre su elección?

Sus profesores.
Sus amigos.
Sus conocidos sociólogos.
Sus familiares.
Sus lecturas sociológicas.
Nadie.

97

En un cuestionario se han categorizado de la siguiente forma estas preguntas:

¿Cuál es la religión que profesa?

Islamismo.
Cristianismo.
Budismo.
Católico practicante.
Anglicano.

¿Dónde vive actualmente?

En Barcelona.
En Castilla.
En Cataluña.
En Madrid.

En el campo.
En Andalucía.
En la ciudad.

Se pide indicar razonadamente si están bien o mal categorizadas estas preguntas recordando que las categorías deben ser exhaustivas, que se excluyan mutuamente y que sólo presenten una dimensión.

RESPUESTA

Las dos preguntas están mal categorizadas en los tres aspectos indicados. La primera no es exhaustiva porque hay muchas otras religiones además de las indicadas; la categoría cristianismo y anglicano no se excluyen mutuamente; y, en fin, la categoría católico practicante presenta, además de la dimensión pedida, religión, la dimensión de su práctica.

La segunda no es exhaustiva, porque hay otros muchos sitios donde se puede vivir en España y también en el extranjero. Las categorías 1 y 3 y 2 y 4 no se excluyen mutuamente, porque el que viva en Barcelona y Madrid, puede contestar con todo derecho que también viven en Cataluña y Castilla. Las categorías ciudad y campo pertenecen a otra dimensión, de la geográfica seguida en las restantes categorías. Uno que viva en Barcelona, lo mismo puede contestar que vive en la ciudad que en Barcelona.

98

Se está realizando una investigación sobre el alcance y efectos de la publicidad realizada para difusión del tipo de coche deportivo "Relámpago" por medio de la prensa diaria, de carteles y de la radio.

Se pide la formulación en relación al anuncio en dichos medios de una pregunta de control con el fin de asegurarse de la veracidad y fiabilidad de las respuestas del encuestado en el cuestionario. Como es sabido, para obtener esta prueba es conveniente incluir algunas preguntas deliberadamente capciosas, falsas o engañosas, en algún punto.

RESPUESTA

En el caso propuesto, la pregunta pedida se podría formular así:

¿En qué medios publicitarios ha visto usted anunciado el modelo de automóvil "Relámpago"?

En la radio.
En la prensa diaria.
En la televisión.
En una revista.
En carteles.

Nótese que se incluyen deliberadamente dos medios en los que, según el enunciado, no se ha anunciado el modelo de coche indicado en la televisión y las revistas. Es claro que se puede dudar de la sinceridad y fiabilidad de las respuestas de todos aquellos que contesten que lo han visto anunciado en estos medios.

99

En un cuestionario, en el que uno de los aspectos sobre los que se desea recoger información es la intención de cambio de ocupación del encuestado, se pretende obtener alguna prueba de la consistencia o solidez de las opiniones que emitan los encuestados al respecto.

Se pide indicar la forma de verificar esta prueba y redactar la pregunta o preguntas que se incluirían en el cuestionario con este fin.

RESPUESTA

El procedimiento que se suele utilizar al efecto es introducir en distintos lugares del cuestionario dos preguntas similares. Si la contestación a ellas del mismo individuo no son iguales, demuestra la poca solidez de sus opiniones o bien la falta de interés y atención al contestar las preguntas.

Suponiendo que el cuestionario tenga 50 ítems, se podrían colocar en los lugares 14 y 38, por ejemplo, las siguientes preguntas.

¿Piensa usted cambiar de ocupación en el futuro?

Sí.
No.
No sé.

¿A qué ocupación piensa usted dedicarse en el futuro?

100

Indicar razonadamente si está bien formulada esta pregunta, sin ninguna otra previa, en un cuestionario.

¿Está usted satisfecho con su nuevo destino?

Sí.
No.

RESPUESTA

No está bien formulada porque da por supuesto la existencia de un nuevo destino y porque la respuesta negativa puede significar tanto que no está satisfecho como que no ha tenido un nuevo destino. Debería ir precedida de otra tipo filtro:

¿Ha tenido usted un nuevo destino dentro de los últimos doce meses?

Sí.
No.

101

Indicar razonadamente si en un cuestionario sobre actitudes hacia el pueblo norteamericano están bien formuladas las dos preguntas que siguen, que aparecen en él una inmediatamente a continuación de la otra.

¿Qué opina usted sobre la intervención norteamericana en el Vietnam?

¿Cuál es su juicio sobre el pueblo norteamericano?

RESPUESTA

Están mal formuladas, porque se puede dar en ellas lo que se llama "contagio de preguntas", o sea, el hecho de que la contestación a una de ellas influya en la opinión que se exprese en relación a la otra. En este caso concreto es posible que en algunos encuestados el sentimiento que despierte la intervención norteamericana en la guerra del Vietnam repercuta en la opinión que exprese en la contestación al juicio que le merece el pueblo norteamericano. Esta influencia por el hecho de estar juntas desaparecería, colocándolas alejadas en el cuestionario.

102

Indicar razonadamente si está bien o mal formulada la pregunta:

¿Cuál es su sueldo?

RESPUESTA

No está bien precisada, porque existe un sueldo íntegro y otro líquido, sensiblemente diferentes entre sí. Por tanto, es imprescindible precisar en la pregunta a cuál de ellos se refiere. Por otra parte, como hoy es muy frecuente percibir más de un sueldo, es necesario también señalar en la pregunta si se trata del sueldo que gana por todos sus empleos, o sólo por su ocupación principal.

La pregunta, por tanto se podría formular así:

¿Cuál es el sueldo neto total que gana, incluidos todos sus empleos?

En todo caso se debe señalar la conveniencia de presentar esta pregunta no abierta como se ha hecho, sino categorizada; no sólo por los motivos indicados que aconsejan la categorización, a ser posible, de las preguntas, sino para facilitar la contestación de este tipo de preguntas enojosas y asegurar, en alguna medida, una mayor veracidad en las contestaciones.

103

Indicar razonadamente si están bien formuladas o no las siguientes preguntas:

¿Qué clase de coche le gusta más?

¿Qué especie de amigos tiene usted?

¿Cuál es la cosa en que se ocupa en la actualidad?

RESPUESTA

No están bien formuladas, porque las palabras clase, especie y cosa son abstractas, generales y demasiado ambiguas, y por tanto se corre el peligro de que cada encuestado las interprete de forma diversa. Por ejemplo, la clase de coche unos la pueden referir a la marca, otros al tamaño, otros al tipo familiar, deportivo, utilitario, etcétera.

104

Señalar si está o no bien formulada la pregunta siguiente, e indicar en su caso la forma que se estime correcta.

¿Cree usted que deben existir sindicatos para los funcionarios públicos y que estos sindicatos deben tener alguna relación con los de los trabajadores?

RESPUESTA

No es adecuada, porque comprende en realidad dos preguntas distintas y uno, por ejemplo, puede ser partidario de la sindicación de los funcionarios públicos, pero no de que éstos sean especiales.

Habría que desdoblarla en dos preguntas separadas. Por ejemplo:

¿Cree usted que deben existir o no sindicatos de funcionarios públicos?.

Sí.
No.

¿A su juicio, éstos deben o no ser completamente indepedientes y separados de los laborales?

Sí.
No.

105

En un cuestionario elaborado para investigar las actitudes sociales se ha formulado la siguiente pregunta:

¿ Sí es usted partidario de la Seguridad Social:

Indicar razonadamente si está o no bien planteada.

RESPUESTA

Se ha de responder que no, sobre todo porque está redactada de forma que sugiere una respuesta afirmativa. Se ha de tener en cuenta, como señala el profesor

Bugeda (1974, pág. 86), que, "en caso de duda o de dificultad de comprensión, todo el mundo tiene la tendencia a contestar de modo afirmativo".

Por ello, en general, las preguntas se deben formular de forma que sugieran igualmente la posibilidad de contestar afirmativa o negativamente.

Por ejemplo, la pregunta del ejercicio se podría formular así:

¿Está Vd. a favor o en contra de la Seguridad Social?

A favor.
En contra.

106

Supuesto que se está investigando la extensión de la ideología integrista, indicar razonadamente si sería aceptable formular en el cuestionario la pregunta:

¿Nos puede decir si es usted o no integrista?

RESPUESTA

No, porque el término integrista está ya ligado a un prejuicio social negativo —no está bien considerado el integrismo socialmente—, y por tanto, lo más fácil es que el encuestado conteste que no. Igualmente ocurre, en sentido favorable o desfavorable, con las palabras de izquierdas —todo el mundo, por lo general, quiere ahora pasar como de izquierdas, y al contrario de derechas—, socialismo, propiedad privada, capitalismo, totalitarismo. Todos estos términos y muchos otros que levantan prejuicios por estar marcados socialmente se deben evitar emplearlos directamente en las preguntas de los cuestionarios.

Lo que hay que hacer en este caso es formular una serie de preguntas relacionadas con el integrismo, que indiquen si el encuestado se adhiere o no a la ideología integrista.

107

Indicar razonadamente si estaría bien formulada una pregunta del tenor siguiente:

¿Puede decirnos por qué no tiene usted automóvil?

RESPUESTA

No, en principio. Como regla general se deben evitar en los cuestionarios los porqués, por las razones siguientes: Primera, dan lugar a una multiplicidad de respuestas que complican mucho el trabajo posterior de codificación. Segunda, no se ajustan a la regla de Bowley indicada, según la cual las preguntas han de estar hechas de tal forma que requieran siempre una respuesta numérica, o simplemente una afirmación o una negación, o bien la elección entre diversas categorías propuestas. Tercera, tampoco serían muy fiables las contestaciones, pues los motivos son frecuentemente complejos y a veces inconscientes.

Lo más acertado sería formular esta pregunta comprendiendo una serie de categorías como posibles respuestas, indicativas de los motivos que se crean más frecuentes e imperantes de no tener automóvil. Por ejemplo:

¿Quiere usted indicarnos por cuál o cuáles de los siguientes motivos no se compra un automóvil?

No me gusta conducir.
La conducción es hoy muy peligrosa.
No tengo dinero suficiente para comprarlo y mantenerlo.
Los coches, sobre todo en las ciudades, plantean más problemas que resuelven.
Se oponen mis familiares.
No es rentable.
Otros. ¿Cuáles?

108

Enjuiciar críticamente si están bien formuladas las siguientes preguntas, dirigidas a matrimonios la primera, y la segunda a jóvenes campesinos:

¿Cuántos hijos nacidos vivos han tenido en su matrimonio?

¿Lee usted mucho?

RESPUESTA

Están mal formuladas. La primera porque antes de hacer esta pregunta es necesario hacer otra para determinar si han tenido o no hijos en el matrimonio, pues, si no es así, no se sabe si los que no contestan es porque no han tenido hijos o porque no quieren contestar a esta pregunta.

La segunda, porque el adjetivo mucho implica un juicio de valor subjetivo y es por ello impreciso. Lo que a unos puede parecer mucho, acaso será para otros sólo bastante, regular o poco.

109

La siguiente pregunta de un cuestionario: ¿Cuáles son las repercusiones socioculturales del uso de los estupefacientes? ¿Se debe considerar bien o mal formulada?

RESPUESTA

No. En primer lugar, porque se estima conveniente con base en la experiencia que las cuestiones se propongan de forma personalizada –¿Qué opina usted de...? ¿Qué piensa de...? ¿A su juicio, cuáles son...?–, ya que se ha comprobado que de este modo las respuestas son más numerosas y exactas.

En segundo lugar, no está redactada –como es una norma general– en un lenguaje familiar, comprensible fácilmente para todo el mundo. Por ejemplo, los términos socioculturales y estupefacientes es fácil que sean poco conocidos de algunos. Por ello se debería redactar en forma similar a la siguiente:

¿En su opinión, cuáles son los efectos en la vida social del uso de las drogas?

110

Exponer razonadamente si están bien formuladas, sin más indicaciones, las preguntas siguientes:

¿Trabaja usted por cuenta propia o ajena?

¿Es usted propietario o arrendatario de su explotación agrícola?

RESPUESTA

No. Porque están formuladas de forma alternativa y se corre el riesgo de que se conteste sí o no, y entonces no se sabe si dice que sí, que es propietario o que sí es arrendatario, o que no es propietario, pero sí arrendador, y al revés, o que no es ninguna de estas dos cosas, y lo mismo en la primera.

Por ello, se debería formular así:

¿Trabaja usted por cuenta propia o ajena?

Por cuenta propia.
Por cuenta ajena.

¿Es usted propietario o arrendatario de su explotación agrícola?

Propietario.
Arrendatario.

111

Indicar razonadamente si está bien formulada la pregunta:

¿No va usted a los toros?

Sí.
No.

RESPUESTA

No, porque además de incitar a contestar no, puede inducir a error a la gente, sobre el sentido del sí. El sí que contesten unos puede significar que "sí va a los toros" y el de otro que "sí, no va a los toros". En general se deben evitar las preguntas expresadas en forma alternativa, como la anterior, y en forma negativa, como la presente.

112

Indicar razonadamente si serían acertadas en un cuestionario las preguntas siguientes:

¿Cuántas veces va usted al cine durante el año?

¿Cuánto se paga en su casa de electricidad al año?

RESPUESTA

No, porque en ambas se le fuerza al encuestado a recordar y a hacer cálculos, y existe el peligro de que conteste con una cifra aproximada que a veces puede ser muy alejada de la realidad.

Por ello, en cuanto a la primera pregunta, se podría categorizar de una forma similar a la siguiente:

¿Con qué frecuencia va usted al cine?

Más de una vez por semana.
Cada semana.
Cada quince días.
Una vez al mes.
Una vez cada dos meses.
Una vez al trimestre.
Tres o menos veces al año.

Respecto a la segunda se podría preguntar:

¿Cuál fue el importe del último recibo de electricidad para uso de su casa que pagó?

¿Qué período comprendía?

113

Exponer cómo preguntar en un cuestionario a los encuestados ¿cuántos oficios o profesiones han tenido en su vida?

RESPUESTA

No sería adecuado preguntárselo así directamente, porque se les obligaría a esforzar la memoria, sobre todo si han desempeñado varias profesiones. Sería preferible facilitarles el recuerdo, procediendo gradualmente, de modo similar al siguiente:

¿Cuál fue su primer oficio o profesión?

¿Cuál fue su oficio o profesión siguiente?

¿Si ha tenido usted otras ocupaciones distintas de la actual, puede indicar cuáles?

¿Cuál es su ocupación actual?

114

Indicar razonadamente si estaría bien formulada en un cuestionario la pregunta:

¿Cuántos aparatos electrodomésticos se poseen?

RESPUESTA

Tampoco sería conveniente hacer esta pregunta directamente, porque obligaríamos al encuestado a forzar la memoria y sería muy probable que se olvidase algunos de los que interesan a la encuesta de modo especial.

Por ello lo procedente sería hacer una lista de los electrodomésticos que se pretende conocer si se poseen o no y formar con ellos una pregunta del tenor siguiente:

¿Puede usted indicarnos de esta lista de aparatos electrodomésticos los que tiene en su casa?

Radio o transistor.
Televisión.
Nevera eléctrica.
Aspirador.
Lavadora.
Fregaplatos, etc.

115

Supóngase que se tiene que investigar la extensión de las prácticas de higiene personal, y las religiosas, ¿cómo se preguntaría a una persona si se lava los dientes diariamente y si va los domingos a misa?

RESPUESTA

Se trata de preguntas embarazosas que los encuestados pueden tener dificultad en contestar, pues les pueden parecer que si contestan que no, pueden quedar al descubierto como desaseados o irreligiosos.

Por eso, en el primer caso se podría preguntar:

Hay personas que tienen una dentadura muy sana y no necesitan cuidársela de modo especial. En cuanto a usted, ¿puede decirnos con qué frecuencia se lava la boca con un antiséptico dental o con cepillo y un dentífrico?

Después de todas las comidas.
Dos veces al día.
Una vez al día.
Un día sí y otro no.
Cada tres días.
Una vez por semana.
Una vez al mes.
Menos de una vez al mes.

En el segundo caso sería conveniente proponerle una serie de actividades frecuentes en los domingos, y entre ellas la misa, y pedirle indicar cuáles de ellas realiza.

De las siguientes actividades, ¿nos podría indicar cuáles realiza usted por lo general todos los domingos?

Pasear con los amigos.
Salir de excursión con la familia.
Ir al futbol.
Ir al cine o al teatro.
Ir a misa.
Hacer deporte.

116

Indíquese cómo preguntar en cuestionario cuánto dinero gana una persona.

RESPUESTA

Se trata de una pregunta si no embarazosa, sí enojosa, pues, por lo común, a nadie le gusta sepan cuánto dinero gana. Además, está sin duda extendido el hábito de disminuir los ingresos para evitar impuestos, gravámenes, compromisos, etc.

Por ello, no se debe preguntar en los cuestionarios de manera directa "¿Cuánto gana usted?", sino hacerlo de alguna forma indirecta, por ejemplo:

He aquí una escala de ingresos mensuales. ¿Podría decirnos en qué grupo de los que comprende se considera usted incluido?

Menos de 30.000 pesetas al mes.
De 30.000 hasta 45.000.
De 45.000 hasta 70.000.
De 70.000 hasta 90.000.
De 90.000 hasta 120.000.
De 120.000 hasta 170.000.
De 170.000 hasta 200.000.
De 200.000 o más al mes.

117

Señalar razonadamente si en los siguientes pares de preguntas iguales, utilizadas en realidad en sendos cuestionarios, redactadas de modo diverso, puede influir en las respuestas obtenidas las diferencias de redacción.

Primer par (propuesto a propósito del tercer mandato presidencial solicitado por Roosevelt).

Redacción A)

¿Es usted partidario de interpretar legamente la Constitución de forma que se impida al Presidente de los Estados Unidos solicitar un tercer mandato?

Redacción B)

¿Es usted partidario de modificar la Constitución de los Estados Unidos de forma que se impida al Presidente solicitar un tercer mandato?

RESPUESTA

1. La diferencia entre ambas preguntas es clara; en la primera se habla sólo de interpretar la Constitución, y en la segunda, de reformarla.

2. Es obvio que la segunda fórmula hubo de encontrar más oposición, como así ocurrió, porque la Constitución en Estados Unidos y en general en todos los países muchos la consideran inalterable, mientras no tendrían ningún reparo que oponer a que se interpretara legalmente en dicho sentido.

Segundo par (propuesto en Estados Unidos en 1940, segunda guerra mundial).

Redacción A)

¿Cree usted que los Estados Unidos deberían hacer más de lo que hacen para ayudar a Inglaterra y Francia?

Redacción B)

¿Cree usted que los Estados Unidos deberían hacer más de lo que hacen para ayudar a Inglaterra y Francia en su lucha contra Hitler?

RESPUESTA

1. La diferencia entre ambas redacciones es simplemente que en la segunda se añade "en su lucha contra Hitler".

2. Ahora bien, Hitler, entonces en los Estados Unidos y hoy también en todo el mundo, tenía un prestigio muy negativo, de modo que su nombre despertaba un sentimiento de oposición y rechazo. Por eso sólo el hecho de incluir esta palabra en el cuestionario puede influir en las respuestas obtenidas, como de hecho influyó en este caso, disminuyendo del 22 al 13 por 100 el número de los que contestaron que los Estados Unidos no deberían hacer más.

Estos ejemplos indican la influencia de la forma de redacción de las preguntas en las respuestas obtenidas. El sociólogo debe tenerlo en cuenta buscando enunciarlas de tal forma que se conozca la realidad lo más objetivamente posible, al contrario de los políticos, que, por ejemplo, en los referendums formulan las preguntas de la forma más ventajosa para sus intereses.

118

En los Estados Unidos se preguntó en un cuestionario a las amas de casa: ¿Cuál es el trabajo casero que teme usted más, el que usted aplaza hasta el máximo? Exponer razonadamente si está bien formulada.

RESPUESTA

Hay que tener mucho cuidado con las explicaciones en las preguntas, pues pueden fácilmente oscurecerlas, y con frecuencia no son necesarias. Así, en este caso, no es necesaria la explicación "el que usted aplaza hasta el máximo". Además, oscurece la pregunta, pues puede haber personas que precisamente porque les repugna, lo ejecutan antes, para quitárselo de encima.

119

En un cuestionario dirigido al público en general se ha formulado la siguiente pregunta:

¿Cómo calificaría usted las relaciones entre personal docente y discente de los organismos de enseñanza superior en los últimos quince años?

Examinar de manera detenida y general si es una pregunta técnicamente admisible, teniendo en cuenta el destino del cuestionario y su utilización, y empleando para efectuar el contraste el elenco de condiciones que deben reunir las preguntas de los cuestionarios que se encuentran en la obra de E. Noelle "Encuestas en la sociedad de masas" (Alianza Editorial, página 71), formado por los siguientes interrogantes:

1. *¿Dificultades de comprensión lingüística?*
2. *¿Pregunta demasiado abstracta?*
3. *¿Se pide demasiado a la capacidad de expresión del entrevistado?*
4. *¿Insuficiente capacidad de observación para responder a la pregunta?*
5. *¿Se pide demasiado a la memoria?*
6. *¿Peligros de la fatiga? ¿Cuestionario demasiado extenso?*
7. *¿Peligro de respuesta por amabilidad?*
8. *¿Negativa por miedo o desconfianza?*
9. *¿Cuestión de prestigio?*
10. *¿Bloqueada la disposición a informar por tema demasiado privado?*
11. *¿Posibilidad de conflicto en el yo ideal?*
12. *¿Se trata de una experiencia inconsciente?*

RESPUESTA

El contraste de la pregunta en cuestión con los interrogantes citados muestra que su lenguaje es poco comprensible, por ejemplo, personal docente y discente; que es demasiado abstracta, pide que se califique el tipo de relaciones; que exige demasiado de la capacidad de expresión del entrevistado, por lo menos una gran parte, dada la generalidad de la encuesta; que existe, sin duda, en muchos de los preguntados insuficiente capacidad de observación respecto al objetivo de la pregunta; que se pide demasiado a la memoria, es muy largo el período de quince años; que puede, en fin, producirse la negativa a contestar por miedo o desconfianza, todo lo cual indica con toda claridad que la pregunta no es apropiada para el cuestionario pretendido.

120

En una investigación realizada en una gran ciudad, se desea saber el medio de locomoción utilizada habitualmente entre el domicilio y el lugar del trabajo. Redáctese un pequeño cuestionario.

RESPUESTA

Sexo:

Varón.
Mujer.

¿Cuál es su estado civil?

Soltero.
Casado.
Viudo.
Otros.

¿Indique, por favor, en qué grupo de edad está comprendido?

De catorce a veinte años.
De veintiuno a treinta.
De treinta y uno a cuarenta y cinco.
De cuarenta y seis a sesenta y cinco.
De sesenta y seis en adelante.
Sin respuesta.

¿Cuál es su profesión?

Obrero sin especializar.
Obrero especializado.
Técnico de empresa.
Empleado doméstico.
Empleado administrativo y funcionario público.
Profesor liberal.
Otros.
Sin respuesta.

¿Utiliza algún medio de transporte para ir a su trabajo?

Sí.
No.
Sin respuesta.

En el caso de que no los utilice, ¿cuál es el motivo de no utilizarlos?

Trabajo cerca de donde vivo.
Me gusta andar.
No tengo suficiente dinero para pagarlo.
Otros.
Sin respuesta.

¿Tiene coche propio?

Sí.
No.
Sin respuesta.

En caso afirmativo, ¿lo utiliza para ir a trabajar?

Siempre.
Frecuentemente.
Raras veces.
Sin respuesta.

En el caso de no utilizarlo nunca, ¿por qué no lo utiliza?

Es muy enojoso circular por Madrid.
No resulta económico.
Se pierde tiempo.
Es un problema el aparcamiento.
Otros.
Sin respuesta.

¿Qué medios de transporte públicos dispone para ir al trabajo? (Señale todos los que sean.)

Metro.
Autobús.
Microbús.
Camionetas.
Ferrocarril.
Autobús de empresa.
Tranvía.
Otros.
Sin respuesta.

¿Cuál de ellos utiliza normalmente?

Metro.
Autobús.
Microbús.
Camionetas.
Ferrocarril.
Autobús de empresa.
Otros.
Sin respuesta.

¿Por qué motivos utiliza normalmente este medio de locomoción?

Rapidez.
Comodidad.
Economía.

Seguridad.
Otros.
Sin respuesta.

121

Dadas las siguientes preguntas para un cuestionario, indicar su tipo

a) Por valor sería tan amable de darnos su opinión sobre la importancia para Vd. de los siguientes atributos de un posible marido:

	Muy importante	*Importante*	*Indiferente*	*Indeseable*
Buena presencia				
Amabilidad				
Riqueza				
Formalidad				
Sentido del humor				
Ambición				
Moralidad				
Inteligencia				
Buena salud				

RESPUESTA

Se trata de una pregunta que se puede calificar de evaluativa y escalar y sobre aspiraciones.

También se le puede atribuír un carácter proyectivo.

Oppenheim, (1966, 82, obra en la que se basa en este ejercicio, incluye este pregunta entre las que el llamado *check-list*, que se podría traducir por lista de comprobación.

b) ¿En qué lugar de los siete puntos de estas escalas situaría Vd. a su madre?

De corazón ardiente	———————	*De corazón frío*
Superprotectora de sus hijos	———————	*Despreocupada de sus hijos*
Paciente	———————	*Impaciente*
Competente	———————	*Incompetente*
Muy igual a Vd.	———————	*Totalmente distinta de Vd.*
Desearía tenerla como amiga	———————	*No desearía tenerla como amiga*
Alegre	———————	*Triste.*

RESPUESTA

En general, es también como la anterior doblemente categorizada, de tipo proyectivo y escalar pero no sobre aspiraciones sino de evaluación o juicio de carácter valorativo. Oppenheim, en la citada obra la llama *rating*, en cuanto atribuye un valor numérico a un juicio determinado.

Entre las preguntas de esta clase también incluya la siguiente:

A continuación se exponen algunas de las razones que los estudiantes exponen como motivo de su insatisfacción con su condición en este país. Indique por favor, la importancia que han tenido en su caso.

	Muy grande	Grande	Alguna	Ninguna
Ausencia de padres, familia.				
Ausencia de amigos intimos.				
Separación de esposa, marido, hijos.				
Pérdida de la lengua propia, libros, revistas, etc.				
Pérdida de festivales, celebraciones, funciones sociales.				
Dificultades alimenticias. Largo periodo de ausencia de su patria, casa.				
Tedio y monotonía de la vida presente.				
Sensación general de depresión psicológica (emocional).				
Otras razones (Especifique, por favor).				

c) A casi todo el mundo, le molesta una vez u otra alguna cosa. Individuos diferentes se molestan por cosas diferentes. De las que siguen, díganos las que le molestan mucho, poco o nada.

	Mucho	*Poco*	*Nada*
No llevarme bien con otro.			
No ser popular.			
Que se rian de mi otros niños.			
Que me llamen ...			
Sentir que me desprecian.			
Que un niño no quiere venir a mi casa.			
Olvidarme de cosas.			

RESPUESTA

Se trata, igualmente, de preguntas doblemente categorizadas, proyectiva escalar y expresiva de los hechos que a cada uno le disgustan. Oppenheim las denomina inventarios.

Otros ejemplos de este tipo de preguntas que formula este autor son los siguientes:

"He aquí un conjunto de afirmaciones sobre los efectos de la TV. Díganos, por favor, si son verdaderas o falsas en su caso.

	Verdad	Falsa
He copiado de la TV la manera de vestir.		
He intentado cocinar algo después de verlo en TV.		
He leído un libro después de oír hablar de él en TV.		
He ido a un museo o galería después de oir hablar. de él en TV		

"Digamos, por favor, cómo se debe calificar las acciones siguientes en una enfermera psiquiátrica:

	Muy mal	Mal	Indiferente	Muy bien	Bien	Depende
Gastar mucho tiempo en hablar con los pacientes.						
Permitir que estén desarreglados.						
Aparentar mucha dedicación y eficiencia.						
Olvidar ponerse un uniforme limpio.						
. . . . .						

d) "Existen muchos remedios bien conocidos que pueden usarse para remediar los distintos tipos de dolencias y enfermedades. Diganos por favor cuales son mejeores para cada una de las dolencias que se indican.

	Gárgaras	*Inhalantes*	*Gotas*	*Jarabes*	*Tabletas*
Resfriado.					
Catarro.					
Bronquitis.					
Pérdida de la voz.					
Insomnio.					
Indigestión. *Asma.*					
. . . . .					

RESPUESTA

Se trata como siempre de una respuesta doblemente categorizada en este caso no escalar, referentes a hecho, y que también se puede considerar como proyectiva e indirecta. Oppenheim las llama *grids*, rejillas.

En general, todos estos tipos de preguntas que abarca este ejercicio pueden tener múltiples usos en cuestionarios, como señala Oppenheim en el lugar citado y servir para probar hipótesis diversas. Cuestiones tabú, inadmisibles y formas antisociales de conducta pueden ser insertadas fácilmente en ellas, sin que llamen la atención y se suelen responder sin los recelos y prejuicios que den lugar si se hacen preguntas independientes. Estimulan la expresión espontánea de juicios y opiniones y conductas muy distintas, dado su carácter indirecto, encubierto por actividades de muy diversa significación.

18. ENTREVISTA

NOCION E IMPORTANCIA

En toda observación social se pueden distinguir tres elementos: la situación o realidad a observar, el instrumento de observación y el acto de observación.

Con relación a ellos la entrevista ofrece la particularidad de presentar características peculiares. La realidad a observar son datos objetivos (características físicas, económicas, profesionales, sociales, etc.) y subjetivos (ideas, creencias, opiniones, actitudes, sentimientos, conductas) del encuestado.

El instrumento de observación es fundamentalmente el cuestionario, como en las encuestas por cuestionario escrito, pero con importantes particularidades con relación a éste.

La primera es que forma parte del instrumento de observación o cuestionario de forma activa el observador o entrevistador. Es este quien formula las preguntas y anota las respuestas, a diferencia del cuestionario escrito por el mismo encuestado.

La segunda que este cuestionario no es siempre fijo y estandarizado sino que puede consistir a veces, según los tipos de entrevista, en el simple señalamiento de unos puntos o temas de la entrevista.

En cuanto al acto de observación, la entrevista presenta la notable particularidad de que, el hecho mismo de la observación o entrevista, da lugar a que se produzca necesariamente una relación social entre entrevistador y entrevistado, con la consiguiente interacción o influjo social recíproco entre ambos. Esta relación social también existe en la observación participante, pero en ella es genérica en relación a todo el grupo, mientras que la entrevista es eminentemente concreta, personal, directa e inmediata.

Estas peculiaridades de la entrevista alteran y modifican la inmediatez y pureza necesarias en la observación científica.

A pesar de ello, la entrevista es uno de los procedimientos de observación principales en las ciencias sociales.

En estas ciencias hasta ahora no existe ningún procedimiento para observar directamente los datos subjetivos de los miembros de la sociedad. Es obligado recurrir a la entrevista o algún procedimiento similar también indirecto.

Por otra parte, la entrevista representa un procedimiento cómodo y barato para obtener datos objetivos de dichos miembros. Se podrían comprobar directamente pero con mucho más esfuerzo, trabajo y coste.

De hecho, la entrevista es la técnica más usada hasta ahora en las investigaciones sociales y la que sin duda ha contribuido más a los logros de la sociología empírica, como señala Sheuch (1974, p. 266).

LA INTERACCION SOCIAL EN LA ENTREVISTA

Las relaciones sociales no son algo que se produce "ex-novo" sin ningún precedente. Antes al contrario, cada uno en nuestras relaciones sociales, al entrar en contacto con los demás, se forma una imagen personal y social de individuo con quien se relaciona. Se hace así de acuerdo con su presencia y apariencia y la experiencia que tenemos de contactos con personas del mismo tipo y grupo social.

Esta imagen que cada uno se forma, nos induce a hacernos determinadas expectativas y a esperar por tanto un determinado comportamiento de nuestro interlocutor, y nos lleva a adoptar por nuestra parte una actitud consecuente con la imagen formada y las expectativas creadas.

La entrevista, a pesar de su finalidad científica no deja de suponer una relación social y por tanto también actúa en ella el mecanismo social descrito. Este mecanismo puede comprometer la finalidad de la entrevista que es la de obtener respuestas personales, no sugeridas y veraces, que se ajustan a la realidad del entrevistado.

En concreto, por ejemplo, el entrevistador tenderá normalmente a suponer en el encuestado unas opiniones y actitudes, en particular y en su conjunto estructural, similares a las de otras personas del mismo tipo y clase social que haya entrevistado e interpretará en este sentido las contestaciones recibidas.

Además, como ocupa en la entrevista una cierta posición preeminente, es posible que en algún modo imponga al encuestado sus propias opiniones y actitudes.

Por su parte, el encuestado tenderá a desplegar consciente o inconscientemente los mecanismos psicológicos de defensa que está acostumbrado a poner en juego en su relación con los demás.

Asimismo, identificará al entrevistador como miembro de un determinado grupo social. En consecuencia, se inclinará a adoptar una posición ante él, consonante con la experiencia que tenga de su trato anterior con personas de dicho grupo y con los estereotipos existentes sobre ella en el grupo a que pertenece.

Consecuencia de todo ello, es que el equipo investigador y los entrevistadores han de tener en cuenta esta interacción para neutralizarla en lo posible en cada caso concreto, e interpretar según ellas los resultados de la encuesta.

CLASES

En el Diccionario de Sociología de Fairchild se define la entrevista como "la obtención de información mediante una conversación de naturaleza profesional".

Esta definición comprende tanto la llamada entrevista estructurada, formal o con cuestionario, como los diferentes tipos de entrevista sin cuestionario o no estructuradas.

En las segundas no existe un cuestionario al que se tenga que ajustar estrictamente el entrevistador. A este solamente se le indica, además del objeto y fin de la investigación, los diversos puntos sobre los que interesa obtenga información de los entrevistados. Por tanto, queda en todo caso a su arbitrio el número y tipo de preguntas a realizar y el orden y modo de formularlas.

Este tipo de entrevista se emplea, sobre todo, en los estudios exploratorios, previos a investigaciones proyectadas, para obtener un conocimiento básico, del que se carece, de la población y el campo de la investigación, así como cuando la investigación se refiere a un tema difícil y escabroso o el universo es muy heterogéneo. En estos casos no es conveniente aplicar un cuestionario rígido, sino dejar en libertad de acción al entrevistador para actuar según las circunstancias de cada entrevista. (Ver ejercicio 129)

Dentro del grupo de entrevistas no estructuradas se pueden destacar: la entrevista clínica, la profunda y la "focal".

La primera se utiliza en psicoanálisis y en psicoterapia. En ella ni siquiera están definidos los puntos a tocar. Al contrario se trata de dejar plena libertad al entrevistado en la expresión de sus vivencias y recuerdos. Lo que se pretende es que éste adquiera conciencia de sus complejos y se libere de ellos.

La entrevista profunda se orienta a investigar en profundidad una determinada cuestión del entrevistado. Por ello el entrevistador debe procurar que las manifestaciones del entrevistado giren alrededor de dicha cuestión. Hay pues una mayor intervención del entrevistador que en la clínica.

La técnica de la entrevista focal "focused interview" ha sido desarrollada por R. K. Merton. Su objeto o tema es una experiencia muy concreta. Por eso se llama focal.

Las personas a quienes se entrevista son las que han participado en una situación específica que se quiere investigar. Por ejemplo, han visto un espectáculo determinado, han leído una obra, han vivido una aventura.

Estos diversos tipos de entrevista se pueden combinar entre sí. En este sentido Lazarsfeld y Merton en la obra "Continuities in Social Research" recomiendan esta técnica:

1. Fase de exploración: conversación centrada o de respuestas libres de una muestra restringida que permita descubrir los aspectos del problema y elaborar un cuestionario.

2. Aplicación del cuestionario a toda la muestra.

3. Retorno a la muestra restringida, con conversación centrada para profundizar los puntos más significativos, revelados por el análisis de resultados del cuestionario.

Según la forma de su realización, se pueden distinguir las entrevistas en forma personal y por teléfono. El contacto entre entrevistador y entrevistado tiene lugar en la primera de forma directa e inmediata, cara a cara, mientras que en la segunda, dicho contacto se realiza a través del teléfono.

En la actualidad, la entrevista personal se enfrenta con graves problemas, dado el aumento continuo de los costes de personal y desplazamientos que exige, y el creciente número de entrevistas que no se pueden realizar por la progresiva dificultad de acceder a las viviendas particulares y la renuncia de la gente a admitir extraños en su casa.

Ante esta situación, existe hoy, especialmente en los Estados Unidos, una tendencia a sustituir la entrevista personal por la telefónica. De este modo, se logra una sensible reducción de los costes e, incluso, a veces, tasas de respuestas más altas.

Como es obvio, la encuesta por teléfono exige el empleo de procedimientos especiales de elección de las muestras, que se basan, bien en las listas de números de las guías telefónicas, o en la generación al azar de números para formar los números de teléfono de la muestra a encuestar.

Entre los factores que están contribuyendo al uso creciente de las encuestas por teléfono se encuentran los siguientes (Frey, 1983):

1. La extensión masiva del servicio telefónico. Hoy, en Estados Unidos, el 98 % de las casas tienen teléfono.

2. La existencia de un amplio espectro de investigaciones y estudios sobre la realización de las encuestas por teléfono.

3. La decreciente aceptación, por los motivos indicados, de los métodos tradicionales de las entrevistas cara a cara.

4. Los avances en la tecnología del teléfono y en la práctica de la encuesta por teléfono. El WATS (Wide-area telephone service) permite realizar llamadas a larga distancia con facilidad relativa y a un coste dos tercios inferior al normal. La técnica CATI (entrevista por teléfono asistida por computador) consiste en la conexión de los teléfonos a un ordenador, lo que capacita para el inmediato proceso de las respuestas al cuestionario. Así, en una noche se ha podido disponer de los datos de una encuesta realizada sobre un acontecimiento ocurrido el día anterior.

Por último, otro tipo interesante de entrevista es la de grupo. Se reune a un grupo de personas, de 6 a 12, para que discutan abiertamente alrededor de un tema que les propone el investigador. Este actúa de moderador y puede introducir determinados estímulos para ver qué reacción provocan.

Se puede considerar como una modalidad de esta entrevista, el llamado grupo de discusión, estudiado en España por el Profesor Ibáñez Alonso en su conocida obra "Más allá de la sociología. El grupo de discusión. Técnica y crítica", Siglo XXI, 1979. Los miembros del grupo se eligen de modo que sean representativos de la población que se quiere estudiar. El investigador plantea genéricamente el tema sin participar en su discusión a fin de que el grupo cree su propio discurso de acuerdo con sus preocupaciones e intereses. El análisis se efectúa con base en la transcripción literal de las intervenciones recogidas en grabadora.

VENTAJAS DE LA ENTREVISTA

La encuesta por cuestionario simple presenta con relación a la efectuada por medio de entrevista las ventajas principales siguientes:

1. Su menor coste y exigencia de personal.

2. Evita la diversa influencia que ejercen los entrevistadores en las respuestas de los entrevistados.

3. Asegura de modo más perfecto el carácter anónimo de las respuestas y de la encuesta.

A su vez, la entrevista, por su parte, presenta, respecto al cuestionario simple, las ventajas siguientes:

1. Posibilidad de aplicación a personas analfabetas o que leen y escriben con dificultad.

2. La mayor importancia que los encuestados suelen conceder a la entrevista con relación a la frecuentemente escasa que conceden a los cuestionarios simples.

3. Es más seguro en ella obtener la cooperación del encuestado, lo que hace que sea mucho más elevado, por lo general, el porcentaje de cuestionarios simples no devueltos o no cumplimentados que el de entrevistas fallidas.

4. La entrevista permite, sobre todo, lograr una comprensión de las condiciones psicológicas y ambientales del encuestado en la entrevista, y de su intención y disposición de ánimo en la contestación de las preguntas, así como aclarar el sentido de éstas cuando no sean suficientemente claras.

5. La entrevista permite obtener una información más completa, profunda y rica, sobre todo en cuestiones comprometidas. (Ver ejercicio 128)

CONDICIONES FORMALES DE LA ENTREVISTA

La naturaleza de observación indirecta que tiene la entrevista y la interacción social que se produce durante la misma entre entrevistador y entrevistado, hace que la entrevista sea una técnica que, en mayor grado, sin duda, que las demás técnicas de investigación social, plantea problemas y dificultades para lograr su objetivo, obtener respuestas válidas y veraces.

Por ello, la realización de la entrevista requiere el mayor cuidado y atención y una escrupulosa programación de su desarrollo y ejecución por parte de los entrevistadores.

A este efecto, los manuales de investigación social y las obras sobre la entrevista suelen contener normas que orientan sobre la preparación y realización de la entrevista.

Estas normas se pueden dividir distinguiendo las dos fases de la entrevista: la preparación y la ejecución.

PREPARACION DE LA ENTREVISTA

Una vez elegida la muestra y seleccionadas las personas que han de ser entrevistadas, conviene presentar al encuestador mediante una nota que anuncie su visita y explique los motivos de la entrevista.

La entrevista se debe realizar en un tiempo oportuno, para lo cual es importante conocer la distribución del tiempo de los encuestados. Si es posible, lo ideal será obtener una cita de antemano.

Los entrevistadores deben tener un conocimiento previo del campo social en que van a actuar. De este modo se facilitará el ajuste entre entrevistador y entrevistado.

Un contacto previo con líderes del grupo social a encuestar para darles a conocer los motivos y objetivos de las entrevistas, puede facilitar el éxito de las mismas, si el líder consiente en apoyarlas.

El entrevistador debe ser preparado especificamente para su tarea, mediante un cursillo de capacitación en el que obtenga un conocimiento básico de la investigación social y una información sobre la encuesta y de sus objetivos. Además es corriente entregarles unas instrucciones con las normas a que se deben ajustar en las entrevistas.

EJECUCION DE LA ENTREVISTA

En la ejecución de la entrevista se pueden distinguir los siguientes aspectos: ambiente de la entrevista; contacto inicial entre entrevistador y entrevistado; formulación de las preguntas; anotación de las respuestas y terminación.

Ambiente de la entrevista.— Si el entrevistador puede escoger el lugar de la entrevista, éste debe ser agradable y acogedor, de modo que facilite la comunicación.

Sobre todo el lugar debe estar a cubierto de miradas y oídos indiscretos. Si faltan estas condiciones, el sujeto puede sentirse inhibido en sus respuestas y puede resultar violado el secreto de la entrevista.

La posición de los participantes en la entrevista debe ser cómoda, por ejemplo, sentados y tomando un refresco o un café.

En el contacto inicial, el entrevistador debe causar una primera impresión a ser posible agradable. Para ello deberá ser y mostrarse educado, simpático y atractivo.

Para que el entrevistado coopere adecuadamente a la entrevista es importante que conozca y comprenda el objeto de la entrevista, lo que se pide de él, y los motivos que justifican la entrevista y la hacen de interés para el encuestado y la sociedad.

Por ello en la toma de contacto, el entrevistador, después de presentarse, deberá explicar claramente estos extremos al encuestado.

En este primer contacto se ha de destacar especialmente el carácter estrictamente confidencial y el anonimato de la información. (Ver ejercicio 122)

Formulación de las preguntas.— El encuestador no tiene que dar la impresión de que la entrevista es un exámen o interrogatorio. Por ello tanto en sus preguntas como en las respuestas debe evitar todo aquello que implique crítica, sorpresa, aprobación o desaprobación en sus palabras o en sus gestos.

Si bien el encuestador debe tener en la mano el cuestionario y echarle una ojeada antes de hacer cada pregunta, ésta ha de ser formulada en tono de voz natural y de conversación, evitando el tono de lectura y centrando su atención en el encuestado y no en el cuestionario.

Todas las personas deben ser interrogadas sin que se introduzcan cambios en el enunciado de las preguntas, para evitar la influencia de las opiniones personales del encuestador y la posible variación de significaciones debida al cambio de palabras o giros.

Las preguntas deben ser formuladas en el mismo orden en que figuran en el cuestionario, pues éste se halla dispuesto de modo que se evite el contagio o influencia de unas preguntas sobre otras.

Es conveniente utilizar frases de transición al pasar de un tema a otro, "veamos ahora", "le parece que sigamos con..."

Es necesario que el entrevistado vea que interesa y se da importancia a lo que él dice. Para ello pueden usarse las expresiones que en el trato social son frecuentes, tales como "bien" "claro" "si" "continue" "le escucho", y hacer algún comentario que no exprese aprobación ni desaprobación a las opiniones del encuestado.

Con el fin de hacer las preguntas lo más espontáneas posible, es preciso que las preguntas se sucedan con cierta rapidez, no dejando descanso entre pregunta y pregunta.

El entrevistador va a preguntar y no responder. Cuando se solicite su opinión personal, debe excusarse cortésmente.

No hay ni que decir que ninguna clase de discusiones es admisible. (Ver ejercicios 123, 124, 125, 126, 130).

Anotación de las respuestas.— La experiencia demuestra que la anotación posterior a la entrevista presenta dos inconvenientes: los límites de la memoria humana que no puede retener con fidelidad toda la información y la distorsión que se produce por causa de los elementos subjetivos que se proyectan en la reproducción de la entrevista.

La anotación directa mientras se desarrolla la entrevista es lo más recomendable para recoger con fidelidad y veracidad la información que proporciona la persona entrevistada. Mejor aún, si puede contarse con el auxilio de un grabador, pidiendo consentimiento al interrogado.

Terminación de la entrevista. Se debe distinguir según se trate de un sólo encuentro o de una serie de ellos. En todos los casos, el interrogatorio se debe terminar en un clima de cordialidad. Cuando la índole de la investigación requiere posteriores entrevistas, debe dejarse la "puerta abierta" para los próximos encuentros.

En el caso de que de modo claro el entrevistador se de cuenta de que el entrevistado le está mintiendo descaradamente o le está tomando el pelo, porque considera en broma la entrevista, deberá hacerle ver que aquel es su trabajo con el que se gana la vida y que no puede perder el tiempo. Le rogará por consiguiente que tome la entrevista en serio. Si no hay rectificación patente, el entrevistador debe terminar la entrevista, no olvidándose de consignar el estado de ánimo del encuestado.

EJERCICIOS

122

En un trabajo de campo sobre actitud ante los valores sociales, religiosos y políticos por el sistema de entrevista, se observa que algunas de ellas han sido efectuadas:

a) En la calle de un día frío y lluvioso.

b) Otra a un obrero que ocasionalmente se encontraba vestido en traje de faena en un establecimiento lujoso lleno de gente elegante.

c) Otra estando cerca personas distintas del entrevistado, que acaso podían enterarse de sus respuestas.

d) Otra en la que se indica que el entrevistado se mantuvo en todo momento frío, distante, mostrando claros deseos de librarse de aquello.

e) Otra, en fin, a una persona visiblemente excitada y nerviosa por un altercado que había tenido poco antes.

Se pide: exponer razonadamente si se pueden considerar válidas o no estas entrevistas.

RESPUESTA

Ninguna de ellas se puede admitir. La a) y b) porque van contra la regla de procurar que el entrevistado se encuentre en una situación de comodidad, tanto física como ambiental, que faltan en este caso por las condiciones climatológicas adversas y el ambiente extraño en que se halla el obrero en el segundo caso.

La c) porque se contraviene en ella el principio básico del secreto de las manifestaciones del entrevistado. La d) porque la entrevista requiere que el entrevistado se abra a la encuesta y que se establezca una cierta relación de confianza y comunicación entre el entrevistador y el entrevistado, requisito que es evidente falta en este caso. Por último la e) porque para que las contestaciones del entrevistado puedan considerarse válidas, se debe encontrar en un estado psíquico normal, lo que no ocurre en este caso, como tampoco si estuviera ebrio o drogado.

123

Un entrevistador ya antiguo y experimentado en una encuesta celebrada se ha procurado adaptar a la distinta personalidad de los entrevistados, mostrando un talante distinto según ella, e incluso se ha permitido, basado en su experiencia, alterar en este sentido el orden y contenido de las preguntas, mientras que otro entrevistador, nuevo y principiante, se ha esforzado en permanecer siempre igual, cualquiera que fuese el carácter de los entrevistados y se ha ajustado estrictamente al cuestionario entregado.

Exponer razonadamente cuál de ellos ha actuado mejor.

RESPUESTA

Ha obrado correctamente el segundo, y no el primero. La conducta de éste ha invalidado las entrevistas que ha realizado. Aparte de la modificación en las respuestas que supone las alteraciones en la formulación de las preguntas, su diversa actitud con cada entrevistado implica una influencia distinta en cada uno de ellos.

En las encuestas sociológicas se busca conocer la opinión del grupo, y no las de cada individuo en particular. Por ello las entrevistas han de realizarse por parte del entrevistador de la manera más uniforme posible para que los resultados obtenidos sean comparables y asimilables a los de otras entrevistas, obteniendo así la opinión del grupo.

124

En una entrevista en la que el encuestado ha contestado con respuestas vagas y evasivas: depende, no sé, en parte sí y en parte no..., y en otras ha mostrado dificultad en contestar, el entrevistador se ha limitado a consignar sin más las respuestas y la actitud del encuestado, mientras que en otra entrevista en la que el entrevistado se ha conducido de forma similar el entrevistador ha procurado ayudarle y lograr que precisase sus respuestas con preguntas de ayuda y estímulo.

Indicar razonadamente cuál de los dos entrevistados ha actuado mejor.

RESPUESTA

El segundo, porque, aunque el entrevistado ha de cuidar sobre todo no influir en el sentido de las respuestas del entrevistado, esto no quiere decir que no deba ayudarle si encuentra dificultades de comprensión de una pregunta ni que no deba procurar obtener respuestas claras y definidas en lugar de vagas e imprecisas, que equivalen a no respuestas.

125

Se han efectuado dos entrevistas, una despaciosamente, dando mucho tiempo para responder a las preguntas e incluso intercalando comentarios ajenos a ella, y otra a un ritmo relativamente rápido y sin interrupciones. ¿Cuál de las dos se debe considerar mejor realizada?

RESPUESTA

La segunda. Primero, porque las contestaciones rápidas serán más espontáneas y, por tanto, más reales que aquellas otras en las que se ha dejado pensar mucho al individuo lo que debe contestar. Segundo, porque se ha comprobado que el ritmo de preguntas rápido hace que desaparezcan las barreras e inhibiciones que en el psiquismo del individuo se oponen a la expresión libre y sin cortapisas de sus ideas, sentimientos y actos.

126

Se han realizado dos entrevistas. En una de ellas el entrevistador ha ayudado al entrevistado en aclaraciones y ejemplos y se ha mostrado expresivo mostrando a veces asombro y extrañeza e incluso su aprobación o desaprobación a los informes proporcionados por aquél.

En la otra, el entrevistador se ha limitado a hacer las preguntas sin poner ejemplos de ninguna clase, ni hacer aclaraciones, salvo en casos imprescindibles, cuidando mucho de que sus gestos, palabras y ademanes no manifestasen ningún signo de la impresión que le producían las respuestas. Indicar razonadamente cuál de las dos es válida.

RESPUESTA

La segunda. En las entrevistas y encuestas se buscan las opiniones propias y personales del entrevistado; por ello, una de sus reglas más fundamentales es que el entrevistador no influya en dichas opiniones, por lo menos conscientemente. Toda entrevista es una interacción social en la que hay una influencia recíproca entre las dos partes que intervienen en ella, quiérase o no. Esto es inevitable, y es uno de los

inconvenientes de la entrevista. Pero es inadmisible que el entrevistador ponga ejemplos, haga aclaraciones no necesarias, exprese agrado o desagrado, formule juicios, porque todas estas actuaciones son influencias conscientes que vician los resultados de la entrevista.

127

Se ha proyectado una encuesta entre obreros sin trabajo para investigar la influencia del paro en su posición social –prestigio, autoridad, respeto– en la familia. En la preparación del guión de la entrevista se ha incluido la siguiente pregunta:

¿Cree usted que su situación de paro ha influido o no en el respeto de sus hijos hacia usted? Sí..... No.....

Indicar razonadamente si es suficiente esta pregunta a los fines pretendidos de la investigación y cómo se deberá completar en su caso.

RESPUESTA

No se debe considerar suficiente, pues es una información muy vaga y general, que no permite juzgar sobre la solidez de los motivos en que se funda, ni tampoco contrastar su exactitud con otras informaciones obtenidas en la entrevista u obserdas. Por ello, esta pregunta necesita ser seguida de otras aclaratorias, que precisen su sentido y alcance.

En el caso de que a la primera conteste afirmativamente, se le podría preguntar:

¿Puede precisar en qué aspectos de la vida familiar ha reconocido que sus hijos le respetan menos?

Por el contrario, si contesta que no, se le puede preguntar:

¿En qué ha notado que sus hijos le respetan igual que antes?

¿Cuáles son los motivos de su opinión?

128

Supuesto que se tiene que realizar una investigación sobre los efectos psico-sociológicos en las familias del reducido espacio de la vivienda referida a un suburbio de una gran ciudad, se pide indicar razonadamente que procedimiento de investigación será más adecuado: el cuestionario o la entrevista. Recordando que para decidir esta cuestión se deberá tener en cuenta:

1. *Las características de la población a encuestar.*
2. *La naturaleza y objetivos de la investigación.*
3. *Los medios de personal, tiempo y económicos de que se dispone. (La entrevista los exige mayores.)*
4. *Que ambos procedimientos presentan ventajas e inconvenientes recogidos en los manuales, cuya trascendencia se deberá tener en cuenta en relación a los fines pretendidos de la investigación.*

RESPUESTA

En este caso, tanto las características de la población a encuestar, habitantes de un suburbio de una gran ciudad, que es probable que en gran parte no sepan leer ni escribir, o lo hagan con dificultad, como la naturaleza ecológica y psico-social de la investigación, en cuanto buscar descubrir los efectos psicosociales de las viviendas de espacios reducidos, sería muy recomendable o necesario elegir la entrevista. La actuación personal de los entrevistadores permitirá un reconocimiento de las condiciones de la vivienda y de sus efectos psicosociales en sus habitantes mucho más exactos.

En general, si se dispone de los medios adecuados, la entrevista se debe aplicar en las investigaciones de este tipo, pues, si se cuenta con entrevistadores buenos conocedores de su oficio, permite una visión personal del ambiente y una penetración psicológica más acentuada que el cuestionario.

Por el contrario, si la materia en estudio es más propiamente sociológica, entonces, si no se pretende una penetración psicológica en las actitudes, tensiones, deseos y aspiraciones de los encuestados, puede ser recomendable el cuestionario, que es más económico y exige menos trabajo y tiempo.

129

Se quiere realizar un trabajo de campo sobre la influencia de las circunstancias sociales en la delincuencia juvenil por un equipo de investigación que no conoce bien este problema. Se pide: indicar razonadamente qué procedimiento se puede utilizar para obtener una primera información empírica sobre la orientación y los problemas básicos que plantea la investigación.

RESPUESTA

En este caso, como en los similares, sería procedente utilizar entrevistas informales o no estructuradas, que se podrían realizar a personas profesionales en relación con la delincuencia juvenil: psicólogos, jefes de establecimientos de corrección, componentes de Tribunales Tutelares de Menores, etc., y a algunos jóvenes delincuentes. En ellas se trataría de descubrir los problemas vivos y reales y los aspectos más importantes en relación a dichos puntos, así como, en el caso de los delincuentes, de profundizar psicológicamente en los aspectos de sus problemas sociales a estudiar.

130

Dado el cuestionario que sigue a cumplimentar por entrevistadores, tomado de la adaptación hecha por Roger Mucchielli en su obra "El Cuestionario en la Encuesta psicosocial" (Madrid. Ibérico Europea de Ediciones, 1974, páginas 174 y ss.) del ejercicio contenido en el libro de Goode y Hatt "Metodología de la Investigación Social" (Méjico. Editorial Trillas), y teniendo en cuenta que los objetivos de la encuesta a la que se refiere el cuestionario eran:

- *Indagar el prestigio asociado a ciertos puestos de trabajo.*
- *La relación de esta asociación en su caso con los estratos sociales de la población estudiada.*
- *El motivo del mayor o menor prestigio de las ocupaciones.*

Se pide en general indicar separadamente para cada pregunta:

1. Su tipo y su función en el cuestionario.

2. Su interés en relación a las distintas hipótesis implícitas en ellas, con indicación explícita de las mismas, y respecto también a los objetivos de la encuesta: conocimiento de la situación ocupacional de los encuestados y de la importancia concedida a las distintas ocupaciones y a la suya y los motivos de ello.

3. Las cualidades o los defectos de su formulación y de su posición en la ordenación general del cuestionario.

Pregunta 1ª. Suponed que un joven eminente os pide vuestra opinión sobre la mejor profesión que podría elegir ¿cuál de ellas (puesto de trabajo) le aconsejaríais:

RESPUESTA

Pregunta abierta que provoca la participación del respondiente. Lo sitúa ante un consejo que ha de dar, por lo cual no puede suscitar controversia ni hostilidad contra la encuesta. Por otra parte, la pregunta orienta la mente hacia el objetivo de la encuesta.

Pregunta 2ª. En vuestra opinión ¿cuál es la única cosa importante que debe preocupar a un joven que ha de escoger su profesión o medio de vida?

RESPUESTA

Análoga a la anterior. Sin embargo, empeña más al entrevistado, siguiendo en buena lógica a la pregunta número 1.

Pregunta 3ª. ¿Qué ha hecho usted la semana última desde el punto de vista de sus ocupaciones?

¿Ha hecho un trabajo retribuido?

¿Ha faltado usted a su trabajo por motivo de una enfermedad, del mal tiempo, de una huelga, de una misión temporal?

¿Ha trabajado usted en su casa?

¿Se ha encontrado usted en formación o en la Escuela?

¿Ha estado usted desligado de toda obligación laboral (retiro, rentista, etcétera)?

¿Ha buscado usted trabajando?

¿Otra actividad...?

RESPUESTA

Esta pregunta es de tipo "filtro", y cumple en cuestionario la función de identificar a los que se ha de hacer las preguntas siguientes 3A y 3B. Se deja además al sujeto libertad para explicarse sobre sus ocupaciones.

Pregunta 3ª. A) (Para los que hayan respondido afirmativamente uno de los tres primeros casos precedentes.)

¿Qué trabajo realiza usted?

RESPUESTA

Esta pregunta afronta el problema central de la encuesta. Las hipótesis implícitas en ésta son: 1. Hay profesiones "consideradas" y otras "poco consideradas". 2. Las personas de diferente nivel social reaccionan de diferente manera ante el prestigio de las profesiones.

Pregunta 3ª B) ¿Ha tenido usted el proyecto, durante los cinco años últimos transcurridos, de cambiar de clase de trabajo?

Sí.
No.
Otra respuesta.

(En caso afirmativo plantea la 3B1 y la 3B2; en caso negativo pasad a la pregunta 4.)

P. 3B1. ¿Qué ocupación, concretamente, tuvo usted el proyecto de aprender?

P. 3B2. ¿Cómo llegó usted a decidirse por tal ocupación?

RESPUESTA

La 3B sigue normalmente a la 3A. En cuanto a las 3B1 y 3B2 son lógicas en relación a la 3B y de gran valor para conocer las aspiraciones en cuanto a profesión o bien las profesiones estimadas entre la población considerada, la primera, y para el estudio de las motivaciones, la segunda.

Pregunta 4ª. Ahora voy a pedirle a usted que evalúe un cierto número de profesiones (se muestra la lista preparada). Por ejemplo, mecánico de ferrocarriles. ¿Cuál es vuestra opinión personal sobre la consideración social de que goza este oficio? Califíquela de 1 a 10, siendo 10 la nota más alta y 1 la más baja. (Sigue una lista de 50 profesiones a evaluar cada una de ellas según una escala de 1 a 10 puntos.)

RESPUESTA

El entrevistado se halla ya ahora "envuelto" en la conversación. En primer lugar ha dado su opinión, después ha dicho algo concreto sobre sí mismo. La parte difícil, o sea, la evaluación, ya se encuentra introducida. Lo ha sido lo antes posible, sin que llegue la fatiga, pero una vez hecha la captación del respondiente.

Pregunta 5ª. Cuando usted dice de una profesión que está "muy bien considerada socialmente", ¿cuál es en su opinión la principal característica que le confiere este valor?

RESPUESTA

Investigación del fundamento del "juicio de prestigio" emitido. La pregunta se encuentra situada en este lugar, porque el entrevistado acaba de hacer 50 juicios de este tipo y la pregunta, muy práctica, extrae la lección de sus juicios espontáneos. Si se hubiera formulado esta pregunta antes que la otra, se hubiesen encontrado dificultades para una contestación.

Pregunta 6ª. Gran número de personas no saben con exactitud lo que es un especialista en física nuclear. ¿En qué consiste, a su juicio, el trabajo de un hombre de esa profesión?

RESPUESTA

Pregunta sin relación con el objeto central. Pregunta de tregua cuya formulación –por su comienzo– hace fácil cualquier respuesta. Su interés procede de su posición en sexto lugar dentro de la ordenación general del cuestionario –interés tópico–. Además sirve como criterio de evaluación del nivel socio-intelectual del entrevistado por comparación con la respuesta a la pregunta 14A.

Pregunta 7ª. En su opinión ¿cuánto tiempo, poco más o menos, debe ir un joven a la escuela para elevarse a la mayor altura posible, más adelante, en el plano profesional?

RESPUESTA

Pregunta introducida para evaluar en qué medida cree la gente que la educación y la formación son necesarias para "triunfar" profesionalmente.

Pregunta 8ª. ¿Me permite usted que le pregunte su edad? (Sitúela entre dos tramos de edad previstos.)

De 20 a 24
De 25 a 29
De 30 a 34
De 35 a 39
De 40 a 44
De 45 a 49
De 50 a 54
De 55 a 59
60 o más.

RESPUESTA

Primera pregunta sobre los datos básicos personales de los entrevistados. Formulación delicada. Finalidad importante, puesto que la hipótesis es que los juicios varían con la edad.

Pregunta 9ª. (A formular solamente a los hombres de menos de cuarenta y cinco años.)

¿Ha servido usted en filas durante el transcurso de la segunda guerra mundial? (El cuestionario fue aplicado en 1947.)

9A (En el caso de contestar afirmativamente a la 9.)

Señale usted el arma:

Ejército de Tierra y Aire.
Marina.
Infantería de Marina.
Guardacostas.

RESPUESTA

Pregunta formulada para verificar la hipótesis según la cual la experiencia militar influye sobre la atribución de prestigio a las profesiones civiles.

Pregunta 10. ¿Quién es el que más gana en su familia?

El entrevistado.
Otro.

10A (En el caso de que sea otro.) ¿Qué clase de ocupación tiene?

RESPUESTA

La pregunta 10 introduce la concerniente al jefe de familia 10A, sin sobresaltar al entrevistado en el caso de que no sea él. La pregunta 10 pretende obtener la información necesaria para ponderar la 3A. Nótese que no se pide la profesión, sino la descripción del trabajo. El nombre de la profesión puede, en ocasiones, inducir a error en la tabulación final.

Pregunta 11. ¿Cuál es (o era) la ocupación principal de su padre?

Pregunta 12. ¿Cuál es (o era) la profesión del padre de éste –o abuelo de usted–?

RESPUESTA

Ambas preguntas tienen interés subsidiario. Consideran la movilidad profesional intergeneracional. Posible comparación con la 3A.

Pregunta 13. ¿Es usted inquilino o propietario de la casa que habita?

Inquilino.
Propietario.
Condiciones especiales.
Una habitación.

P. 13A. (Para los inquilinos.) ¿Puede usted decirme a cuánto asciende mensualmente su alquiler?

P. 13B. (Para los propietarios.) ¿Puede usted decirme el valor en alquiler de la casa sin muebles?

RESPUESTA

Medida de status socioeconómico. La hipótesis implícita es que los diferentes niveles económicos reaccionan de diferente manera ante el prestigio profesional.

Pregunta 14. ¿Se acuerda usted de la última escuela a la que ha asistido?

RESPUESTA

Hipótesis implícita: el nivel de formación es un determinante del juicio de prestigio profesional. Nótese que la pregunta empieza con el nombre de la última escuela a que ha asistido. Buena introducción a la pregunta 14A, porque, al mismo tiempo, orienta suavemente hacia los estudios realizados.

P. 14. ¿Cuál es el último año escolar completo que usted ha hecho?

Terminal de secundaria.
Secundaria no terminal.
Enseñanza superior completa.
Enseñanza superior no terminada.
Primaria completa.
De 5 a 7 años de primaria.
De 1 a 4 años de primaria.
Sin escolaridad.

RESPUESTA

Sirve para conocer el nivel socio-intelectual del entrevistado sin hablar de diplomas, éxitos, ni fracasos y también para establecer la muestra por comparación con los porcentajes totales conocidos de asistencia escolar.

Pregunta 15. ¿Su preferencia le inclinaba hacia el partido demócrata o hacia el republicano, con ocasión de las últimas elecciones presidenciales?

Demócrata.
Republicano.
Ninguno.
No recuerda.

RESPUESTA

Evaluación de la hipotética relación entre actitud con respecto al prestigio profesional y tendencias políticas.

Pregunta 16. ¿Ha votado usted o no (por una u otra razón) en las últimas elecciones?

Votó.
No votó.
Demasiado joven.
No recuerda.

RESPUESTA

Evalúa la relación hipotética entre participación política y el objetivo de la encuesta. Nótese la prudencia de la forma que evita cualquier embarazo en la contestación.

A partir de la pregunta 17 el encuestador responderá por sí mismo en función de lo que haya observado. Termina, pues, en la pregunta 15, dando las gracias.

Pregunta 17. Sexo: Masculino. Femenino.

Pregunta 18. Nivel económico del entrevistado.

A) Superior.
B) Medio superior.
C) Medio inferior.
D) Inferior.

Pregunta 19. Dirección (señas).

Pregunta 20. Fecha y firma del encuestador.

19. ESCALAS EN LAS CIENCIAS SOCIALES

Con relación a la palabra escalas se utilizan en las ciencias sociales tres términos: uno genérico, escalas sin más, y dos específicos, escalas sociométricas y escalas de actitudes.

El primer término es genérico porque es común a todas las ciencias y no solo a las sociales.

Es el estudiado en el primer capítulo de esta misma sección, dedicado a la observación, a la medida y a las escalas.

Una escala en sentido genérico se define, según vimos, como la disposición conjunta de cosas distintas pero comunes en algún aspecto.

Como las ciencias sociales se basan en el descubrimiento y análisis de relaciones entre variables, que están formadas por diversas categorías o dimensiones, se puede decir que operan también con escalas. En efecto, el conjunto de categorías o dimensiones de las variables constituyen de hecho una escala que corresponden a algunos de los tipos mencionados: nominales, ordinales, de intervalo y de razón, según la naturaleza de la variable de que se trate. En esta correspondencia se basa la aplicación de los procedimientos estadísticos de análisis en este campo de las ciencias sociales.

Con el nombre de escalas sociométricas se designa en las ciencias sociales los intrumentos de medida propios de estas ciencias, constituidos por la disposición conjunta en un cuadro o cuestionario de los distintos elementos o aspectos de una característica o actitud, a los que se asigna convencional o empíricamente un valor numérico, lo que permite cuantificar el grado de esa característica o actitud que se da en un objeto o sujeto determinado.

Los distintos tipos de escalas, en sentido genérico, nominal, ordinal, de intervalo y de razón representan diversos niveles de medida.

El nivel más bajo, el menos exacto, corresponde a las escalas nominales, y el más elevado y más exacto a las escalas de intervalo y de razón.

En gran parte los procedimientos de observación estudiados aquí corresponden al nivel de medida más bajo, el nominal.

Frente a ellos, las escalas sociométricas pretenden ser instrumentos de observación y de medida más rigurosos. En ellas se alcanza en todo caso el nivel de medida ordinal con puntuaciones específicas para cada sujeto, que en algunos casos se consideran formadas por intervalos iguales y se tratan como escalas de intervalo.

ESCALAS DE ACTITUDES

Aunque se utilizan escalas sociométricas para medir características diversas, las más conocidas, numerosas y utilizadas son las escalas de actitudes, es decir aquellas que se emplean para medir el grado en que se da una actitud o disposición de ánimo permanente, respecto a cuestiones específicas en un sujeto determinado.

Las escalas de actitud tienen en la actualidad una gran aplicación en las investigaciones sociales, bien de manera independiente, bien formando parte de cuestionarios simples o de entrevistas. Dada su esencial condición numérica, como instrumentos de medida, proporcionan al cuantificar los resultados de las observaciones, una precisión más elevada y una mayor posibilidad de aplicación de las matemáticas a su análisis y estudio, así como hacen más viable el contrastar y comprobar en nuevas investigaciones los resultados anteriores obtenidos.

FUNDAMENTO DE LAS ESCALAS DE ACTITUDES

Desarrollando la noción que poco ha he apuntado, las actitudes se pueden definir como disposiciones permanentes de ánimo formadas por un conjunto de convicciones y sentimientos que llevan al sujeto que las tiene a actuar y expresarse según ellas en sus actos y opiniones.

Según la anterior definición, las actitudes comprenden tres factores:

1. Un factor ideológico formado por ideas y convicciones determinadas.

2. Un factor sentimental que implica simpatía o antipatía hacia cosas o personas.

3. Un factor reactivo que impulsa a pensar y obrar en las distintas situaciones sociales en consonancia con las convicciones y sentimientos profesados.

Las actitudes pues se concretan y expresan en ideas, sentimientos hacia cosas y personas y reacciones o modos de actuación diversos.

Por ello, sus manifestaciones en la vida social son múltiples y se puede decir que tiñen todos los campos de la vida social.

Por ejemplo, supuesto un sujeto que sea progresista esta actitud, se reflejará en sus ideas y actos en todos los campos de la vida social: familiar, educativo, político, económico, religioso, cultural, recreativo, etc.

Todas estas manifestaciones constituyen los elementos o atributos de las actitudes. Si fuera posible concretar y expresar todas ellas en un cuadro, en una serie de proposiciones y reunirlas en un conjunto, este cuadro sería representación perfecta de la actitud en cuestión.

Sin embargo, aunque esto no es posible, si lo es reunir en un cuadro no todas las proposiciones referentes a una actitud, pero si las que se crean más significativas. La escala de actitudes no es otra cosa que este conjunto o cuadro.

Los elementos pues de las escalas de actitudes son las proposiciones, se les suele llamar items, que consisten en expresiones que se refieren a ideas, sentimientos y actos en que se concreta o manifiesta la actitud.

Los items o proposiciones de las escalas, hacen las veces de estímulos ante los cuales los encuestados deben reaccionar de una u otra forma.

Estos elementos, proposiciones o items de las escalas no sólo concretan y especifican las actitudes, también facilitan su observación al subdividir la actitud, genérica y abstracta, en rasgos y atributos concretos.

Pero las proposiciones no son únicamente medios o instrumentos de observación sino también lo son de medición cuantitativa, que es lo propio y distintivo de las escalas. Sin ello no serían en sustancia otra cosa que un cuestionario simple.

Las proposiciones o items de las escalas de actitudes, igualmente que éstas tienen de por sí, carácter cualitativo. Para transformarlas en instrumentos de medida cuantitativa, hay que cuantificarlas.

Se realiza, por ejemplo, bien asignando a cada proposición un valor, en relación a la actitud de que se trate, como en las escalas tipo Thurstone, o bien asignando no a cada pregunta, sino a cada respuesta dicho valor, lo que se realiza en las escalas tipo Lickert.

CONDICIONES DE LAS ESCALAS DE ACTITUDES

Para que las escalas de actitudes sean útiles como instrumentos de medida deben reunir ciertas características, dificiles de reunir, entre las

que destacan la graduación, la discriminatividad, la validez, la seguridad, la unidimensionalidad, y el estar formadas por intervalos iguales.

La graduación supone que las escalas deben comprender todos los grados fundamentales de la actitud de que se trate, desde los elevados a los medios y bajos y no sólo parte de ellos.

La discriminatividad implica que sus items deben tener poder discriminante, deben constituir estímulos capaces de distinguir y dividir, incluso en extremos opuestos, a las personas en relación a la actitud de que se trate.

La validez y seguridad son condiciones que, como hemos visto ya se exigen a todo instrumento de observación. Sin embargo, en las escalas revisten importancia especial, dado su carácter de instrumentos de medida cuantitativa.

La validez es la aptitud real de la escala para medir lo que pretende medir. Exige que los resultados obtenidos con la escala reflejen verdaderamente la extensión real en los sujetos de la actividad investigada.

Por su parte, la seguridad pide que la escala en sucesivas aplicaciones a los mismos grupos y en situaciones similares proporcione resultados iguales o similares.

En cuanto a la unidimensionalidad implica que la escala se refiera a una sola propiedad o actitud. Si hace referencia a dos o más no se podrá conocer en realidad la verdadera significación de los resultados obtenidos.

La sexta y última condición indicada, estar formada por intervalos iguales, en rigor no la cumplen las escalas conocidas hasta ahora. No obstante pretenden reunirla y se consideran en la práctica como tales, las escalas experimentales del tipo Thurstone.

TIPOS DE ESCALAS DE ACTITUDES

Existen diversos criterios de clasificación de las escalas de actitudes; Sin embargo, hay dos que parecen tener mayor interés.

El primero, es un criterio de clasificación antiguo, se podría decir clásico, en cierto modo superado, pero que conserva todavía interés práctico. Es el que divide a las escalas según el procedimiento de su valoración en arbitrarias y experimentales.

Escalas arbitrarias son todas aquellas en las cuales el procedimiento de valoración tiene su origen en el convenio o arbitrio de los constructores de escalas. Por el contrario, en las escalas experimentales, dicho pro-

cedimiento de valoración no es subjetivo, sino resultado de un experimento.

El segundo criterio de clasificación es más moderno y tiene una fundamentación teórica mayor. Se basa en qué estan centradas las escalas de actitudes: los sujetos a través de sus respuestas, los items o proposiciones, o ambas cosas a la vez.

Siguiendo al profesor Bugeda (1974, p. 46 y ss.) en el primer caso, a cada sujeto se le adscribirá una cantidad de la actitud considerada, según su puntuación en las respuestas. La mayor parte de las escalas son de este tipo, como las de Likert, Chapin, Sevell y en general, las arbitrarias. Los estímulos o preguntas son aquí una simple repetición que sirve para ordenar las reacciones de cada elemento medido.

En el segundo caso se trata de medir y ordenar los diversos estímulos empleados o proposiciones, con ayuda de un equipo de jueces. Se les denomina escalas de juicio. Las escalas de Thorndike y de Thurstone son de éste tipo.

En el tercer caso, se trata de escalar tanto los estímulos como los sujetos. Ejemplo: los métodos derivados de los escalogramas de Louis Guttman. En el primer tipo la variabilidad de respuestas o de reacciones al estímulo de los items se atribuye a los sujetos, en el segundo, a las diferencias de los mismos estímulos y en el tercero, "la variabilidad de la reacción ha de explicarse ahora en base de las diferencias tanto entre los sujetos, como en los estímulos".

También existe la distinción entre las escalas o tests de tipo A y los de tipo B. En los primeros, los sujetos a través de sus respuestas se sitúan a lo largo del continuo de la actitud variable. En los segundos, los sujetos sitúan un objeto a lo largo de dicho continuo.

ESCALAS DE ACTITUDES ARBITRARIAS

Entre las escalas de actitudes de esta clase se encuentran las siguientes: escalas de puntos, escalas de clasificación, escalas de intensidad y escalas de proposiciones.

Las primeras consisten en una serie de palabras o proposiciones que se presentan al encuestado para que tache o borre las que le sean desagradables, las que espontáneamente le produzcan un sentimiento de repulsión.

Para puntuar estas escalas se pueden seguir diversos procedimientos. Uno de ellos podría ser asignar a cada individuo en principio un número

de puntos igual al de palabras de la escala. A este número se le restan las palabras que representan una posición negativa con relación a dicha actitud, y se le suman las que representan una posición positiva respecto a la actitud medida. (Ver ejercicio 132)

Escalas de actitudes de clasificación

Las llamadas escalas de clasificación consisten en clasificar por orden de preferencia una serie de proposiciones, palabras, ocupaciones, objetos, etc.

La puntuación en este caso se puede realizar sumando los números de los lugares que ocupan en la clasificación, las proposiciones referentes a la actitud que se quiere medir. (Ver ejercicio 133)

Escalas de actitudes de intensidad

Se trata en ellas de elegir respecto a una pregunta o proposición entre varias respuestas que expresan el grado de aceptación o rechazo de la pregunta en cuestión.

Normalmente las categorías de las preguntas en estas escalas suelen ser cinco o siete, del tipo siguiente:

Muy conforme, conforme, indiferente, disconforme, muy disconforme.

En el caso de siete se añaden a las anteriores en los extremos los de absolutamente conforme y absolutamente disconforme.

La puntuación en estas escalas se establece simplemente sumando al número de orden de las categorías elegidas dentro de cada item.

En este tipo de escalas hay que tener en cuenta que la valoración dada a los items o preguntas de la escala que representen la actitud opuesta a la que se mide se computa inversamente. Es decir, que en ella el absoluto desacuerdo vale 1 punto, en lugar de 7, y en absoluto acuerdo 7 en lugar de 1.

A esta clase de escalas de actitudes pertenecen las del tipo Likert, bastante conocidas y empleadas.

Una objeción importante que se hace a las escalas de Likert es que en ellas se valoran igual todos los items, a pesar de que unos pueden denotar una actitud más intensa que otros.

Escalas de actitudes de proposiciones

Se caracterizan porque constan de una serie de actuaciones favorables y desfavorables, a un grupo social, o en general a un fenómeno de interés social.

Dentro de esta especie de escalas es muy conocida la de Bogardus, utilizada mucho en Norteamérica para medir la actitud hacia los negros, o la distancia racial hacia ellos.

Los items o categorías de esta escala de actitudes, que se pueden utilizar en relación a cualquier grupo étnico o racial, son:

1. Aceptarlos incluso para casamiento.
2. Como amigos regulares.
3. Como compañeros de trabajo.
4. Como vecinos.
5. Como conocidos.
6. Les excluiría de mi vecindario.
7. Les excluiría de mi país.

Similar a la de Bogardus es la de distancia social de Dodd. En ella se propone una serie de grupos sociales, por ejemplo, capitalistas, profesiones liberales, oficinistas, y se pregunta que se señale respecto a cada uno de ellos la conducta que se escogería con más agrado de las siguientes:

1. Si deseara casarme, lo haría gustosamente con uno de ellos.
2. Le invitaría gustosamente a cenar.
3. Prefiero considerarlas como personas a las que se conoce de vista, y con las que se cambian algunas palabras al azar si se las encuentra.
4. No me resulta nada agradable el encuentro con una de estas personas.
5. Me gustaría que se suprimiera a todos estos individuos.

La puntuación en este caso se determina simplemente por el número de orden del item elegido. Así por ejemplo, si alguien respecto a los obreros eligiera el item 1, se le asignaría un 1 de distancia social respecto a ellos.

ESCALAS DE ACTITUDES EXPERIMENTALES

Son, como he indicado, aquellas en que cada item tiene un valor o puntuación establecido experimentalmente.

Las más extendidas de estas escalas son las de Thurstone. Formalmente son similares a las de intensidad, en cuanto consisten en un conjunto de proposiciones referentes a la actitud de que se trata. Sin embargo se diferencian de las escalas de intensidad en que cada item o pregunta tiene asignado un valor, el valor escalar, obtenido experimentalmente. Los valores escalares de los items o preguntas deben cubrir toda la extensión o continuo de estas escalas.

Estos valores escalares son causa de que exista una diferencia o distancia entre los valores escalares de cada pregunta, por lo que se dice que estas escalas tienen distancia y por ello se les llama de intervalo.

Para la aplicación de estas escalas únicamente se pide a los encuestados que indiquen si aceptan o rechazan cada item. La puntuación se determina sumando los valores escalares de las proposiciones que aceptan y hallando su media.

Una de las críticas principales que se hacen a las escalas tipo Thurstone es que la valoración hecha por los jueces puede no ser adecuada para el grupo a que se aplique.

Además, estas escalas de actitudes no discriminan a los sujetos que alcanzan un misma puntuación, cuando los items que aceptan son distintos si bien suman lo mismo.

OTROS TIPOS DE ESCALAS DE ACTITUDES

Entre ellos se pueden destacar: el escalograma de Guttman y el diferenciador semántico.

El primero está formado por una serie de preguntas a las que se contesta simplemente si o no, pero formuladas de tal manera y en tal orden que el contestar sí a una de ellas supone normal y lógicamente contestar sí a todas las siguientes.

El propósito de Guttman fué construir escalas verdaderamente unidimensionales, en las cuales sus preguntas o items formasen una jerarquía rigurosa, de modo que la aceptación de una proposición implicara lógicamente la aceptación de las de nivel jerárquico inferior.

En estas escalas están valoradas tanto las preguntas por su orden jerárquico, como los sujetos, porque su aceptación o rechazo de las proposiciones permite conocer su puesto en la escala.

La puntuación total de los encuestados se determina por el número de items o preguntas a las que han respondido sí. También en esta puntuación se deben tener en cuenta el número de fallos o de veces que los síes o noes rompen el orden jerárquico de la escala. (Ver ejercicios 135, 136)

En el llamado diferenciador semántico, se presentan pares de calificativos opuestos, fuerte-débil, grande-pequeño, etc., dividiendo cada par en una graduación de siete puntos. Se debe a Osgood y es en sí originariamente una técnica semántica para medir el significado que tiene algo para un individuo, aunque se puede utilizar también para medir la actitud hacia ese algo.

Una de las funciones de este tipo de escalas es proporcionar una medida indirecta de los sentimientos subyacentes, incluso inconscientes, hacia un término o palabra.

Se puede aplicar este tipo de escalas a diversos objetos o personas y permite captar la impresión que causan en el encuestado, o sea su imagen. Se deben aplicar comparativamente.

El análisis factorial de una gran cantidad de diferenciales semánticos ha descubierto que con ellos se miden tres dimensiones básicas:

a) El valor (bueno-malo), o la evaluación que hace el sujeto de los entes propuestos.

b) El poder, (fuerte-débil), o la percepción por el sujeto de la potencia del ente en cuestión.

c) La actividad (activo-pesado: rápido-lento), o sea la percepción del dinamismo del ente en cuestión.

Como señala J. Padua (1980, 224), Osgood ha propuesto la escala universal siguiente, que como tal puede ser usada para medir la actitud hacia cualquier ser: Bueno-malo; débil-fuerte; exaltado-moderado; absurdo-significativo; blando-duro: pasivo-activo; positivo-negativo; poderoso-impotente; frío-cálido. (Ver ejercicio 134)

CONSTRUCCION DE LA ESCALA DE LIKERT

En la construcción de las escalas se pueden distinguir, en general, las siguientes operaciones básicas:

1. Búsqueda de las proposiciones, items o estímulos representativos de la actitud o conceptos que se pretenden medir.

2. Prueba de cada uno de los items para contrastar si representan adecuadamente a la actitud.

3. Establecimiento de un procedimiento de valoración de la escala, arbitrario o experimental según el tipo de escala elegido.

4. Prueba de la escala completa para comprobar su validez y seguridad.

Veamos ahora en concreto, la forma de construcción de las escalas de tipo Likert.

La construcción de las escalas de Likert consiste en primer lugar en pensar y formular una serie de proposiciones, bastantes, pueden ser cien o más, referentes a la actitud en cuestión.

En segundo lugar dichos items o proposiciones se someten a la prueba de aplicarlos a un grupo, para eliminar los items inadecuados, que no se ajustan bien a la actitud que se pretende medir.

Se averigua para todos los encuestados la puntuación total y se seleccionan de todos ellos el 25 % que ha obtenido una puntuación superior y el 25 % que la haya obtenido inferior. Se forman así dos grupos, superior e inferior de la escalas. Respecto a estos dos grupos se analiza su forma de contestación a cada proposición. Este análisis se basa en el supuesto de que si se trata de un item que se ajusta a la actitud en cuestión, la puntuación media respecto a este item del grupo superior debe ser significativamente más elevada que la del grupo inferior. Para comprobar esta significatividad de las diferencias se utiliza comunmente la prueba estadística t. (Ver ejercicio 205-C)

Los items o proposiciones en los cuales las diferencias entre los dos grupos no sean significativas se eliminan y sólo se dejan en la escala aquellos en lo que dicha diferencia haya resultado significativa.

También se puede comprobar la consistencia de los items de esta escala correlacionando las puntuaciones de los sujetos en el item X con sus puntuaciones totales en la escala menos la del item X. (Ver ejercicio 145)

CONSTRUCCION DE LAS ESCALAS DE THURSTONE

La construcción de las escalas de Thurstone consiste en síntesis en pensar un grupo elevado de preguntas o proposiciones en relación a la

actitud que se pretende medir, favorables y adversas y neutrales en todos sus grados.

Estas proposiciones se someten a un grupo que hace de jueces y se les pide que señalen la valoración que a su juicio objetivo, no personal, tiene el item o proposición respecto a la actitud en cuestión.

Como existe la posibilidad de que los jueces cumplan mal su tarea, por ello se deben analizar las evaluaciones de cada uno, y eliminarlas en el caso de que su trabajo sea deficiente. Por ejemplo, se pueden eliminar los jueces que den el mismo valor a un número elevado de ítems.

La valoración han de hacerla, según una escala continua de 1 a 11 puntos, en el que el 1 significa el máximo acuerdo y el 11 el máximo desacuerdo con la proposición.

Valorados los items de este modo por los jueces, se halla, para cada uno, la media y la desviación típica de las valoraciones que han dado los jueces.

La media constituye la ponderación o valor escalar que se da al item. En cuanto a la desviación típica, los items en que ésta sobrepasa por ejemplo el valor de dos se eliminan.

En lugar de la media y la desviación típica, se pueden utilizar la mediana y la dispersión intercuartílica, lo que se estima más recomendable, ya que, matemáticamente, en realidad, la escala es más ordinal que de intervalo.

Los items eliminados son, por tanto, aquellos en los que la dispersión de las valoraciones de los jueces es mayor, y por tanto es menor su acuerdo respecto a la significación del item en cuestión respecto a la actitud medida.

Suelen quedar unas treinta proposiciones. Se debe procurar que se distribuyan a lo largo de la escala de modo más uniforme posible y cubiendo todos sus valores, sin huecos, del 1 al 11. (Ver ejercicio 141).

CONSTRUCCION DE LAS ESCALAS DE GUTTMAN

Se empieza igual que en los casos anteriores, formulando un número de preguntas o ítems en relación a la actitud de que se trata que pueden llegar hasta 50.

Este conjunto de preguntas se somete para su prueba a un grupo que debe contestar a todas ellas sí o no.

Con las respuestas dadas a todas las preguntas, se forma una tabla, que se llama escalograma.

En este cuadro consignamos la forma de respuesta de cada sujeto, designamos con una X su aceptación de los items y con O su desacuerdo con ellos.

Seguidamente se computa la puntuación total de cada sujeto del grupo. Después estas puntuaciones se ordenan de la más alta a la más baja. A continuación, con las respuestas dadas a todas las preguntas se forma un cuadro que se llama escalograma. La finalidad de esta operación se dirige a jerarquizar rigurosamente los items. En este escalograma los items se disponen en sentido horizontal y los sujetos en sentido vertical, o sea, se atribuyen a los primeros las columnas y a los segundos las filas de la tabla.

Después por aproximaciones sucesivas se forman las nuevas tablas que sean necesarias, hasta que se consiga ordenar por orden jerárquico las proposiciones y las respuestas. A la vez se van eliminando las preguntas que presentan respuestas muy desviadas que infringen la jerarquización debida a las preguntas.

Se estima que la jerarquización de una escala de Guttman, es buena cuando su coeficiente de reproductibilidad es al menos de 0,90. Algunos autores admiten también a partir del 0,85.

Este coeficiente se calcula según la fórmula siguiente:

$$R = 1 - \frac{E}{Q \times S}$$

En el que E es el número total de errores o fallos. Q el número de preguntas del escalograma y S el número de sujetos.

Para computar según la técnica de Goodenough el número de errores, se establecen, con signos más y menos todos los patrones de respuesta posibles, tanto los que se ajusten a la exigencia de jerarquización de la escala como los que se apartan de ella. Después se buscan en el escalograma los casos de sujetos que corresponden a cada tipo de respuestas no jerarquizadas. Este número multiplicado por el número de errores

que suponga cada patrón de respuesta no jerarquizada y su suma nos da el total de errores.

Este escalograma nos ofrece una jerarquización global de la escala y en relación al mismo se puede realizar una primera estimación de la bondad de esta jerarquización calculando el coeficiente de reproductibilidad.

Sin embargo, la técnica anterior no nos dice nada sobre el grado de jeraquización de las respuestas dadas por el grupo, respecto a cada item por separado.

Para remediar este defecto el mismo Guttmann ideó otra técnica de construcción de esta escala, que recibe el nombre de Cornell, por la Universidad donde fue ensayada.

De acuerdo con ella, en la aplicación de la escala al grupo de prueba, cada sujeto debe contestar a cada item como si fuera una escala de Likert, es decir, según diversos grados de acuerdo, normalmente 5.

A continuación se halla la puntuación total de cada sujeto y se ordenan éstos por orden de puntuación de mayor a menor.

Si hubiera una jerarquización perfecta dentro de cada item, entonces los que tienen una puntuación más elevada en la escala completa, deberían tener en cada item una puntuación no inferior a la que ostentan en el mismo items otros sujetos de puntuación global inferior.

Esto se ve graficamente también mediante la construcción de un escalograma más complicado que el anterior. Si no ocurre así, se hacen igualmente que antes manipulaciones en los grados de los items para mejorar su escalabilidad, a la vez que se eliminan los items muy anómalos.

Para un mayor detalle sobre la construcción de esta escala y de las anteriores, se pueden consultar la obra del Prof. Bugeda, *La medida en las Ciencias Sociales* (1975) y la de Jorge Padua (1980).

CALCULO DE LA SEGURIDAD DE LAS ESCALAS SOCIOMETRICAS

El conocimiento de la validez y seguridad de cualquier instrumento de observación es un requisito necesario para saber si dicho instrumento es fiable. Es decir, si ofrece garantías como tal de proporcionar en su empleo repetido resultados veraces y constantes.

En estas escalas, dado su carácter y su pretensión de constituir instrumentos de medida más exactos y rigurosos, estos requisitos son aún más fundamentales si cabe.

La seguridad de estas escalas consiste, como en los demás instrumentos de medida, en su propiedad de proporcionar resultados similares en sus sucesivas aplicaciones.

De acuerdo con la exposición que de este tema hace el profesor Bugeda (1974, pág. 184 y ss.), para hallar la seguridad de una escala sociométrica existen tres procedimientos básicos:

1. Aplicación repetida de la misma escala al mismo grupo.
2. Aplicación de escalas paralelas o similares al mismo grupo.
3. División por mitad de una escala y aplicación de cada parte al grupo.

Aplicaciones repetidas de la misma escala

Consiste, como se ha indicado, en aplicar la misma escala al mismo grupo en dos ocasiones y correlacionar las dos series de medidas constituidas por los resultados de cada sujeto. La correlación obtenida es el coficiente de seguridad.

Este método es el más lógico, pues es el que se ajusta mejor a la definición de la seguridad de la escala. Sin embargo, presenta graves objeciones prácticas.

Si se aplica la misma escala dos veces al mismo grupo, las variaciones de puntuación no se deberán sólo a los errores anejos a cada aplicación de la escala, sino también a una serie de cambios que se producen inevitablemente en cada individuo.

La práctica adquirida en la contestación de la escala influirá más en unos que en otros y podrá ser una fuente de diversidad de una a otra aplicación de la escala.

Igualmente sucederá con la fatiga, el recuerdo de la primera aplicación, la mejor comprensión de lo que se está haciendo.

Todo ello se traducirá en diferencias de puntuación que no se deben a la escala misma.

Aplicaciones de formas paralelas de la escala

En psicología se suele solucionar el problema anterior, aplicando al mismo grupo formas paralelas, pero diferentes, de un mismo test y correlacionando los resultados.

Sin embargo, aunque en el caso de los tests, sobre todo de cierto tipo, es factible construir tests paralelos, no lo es así en las escalas sociométricas.

En éstas, una escala sociométrica paralela a una dada, es siempre otra escala, que probablemente tendrá un significado distinto y medirá otra dimensión.

División por mitad

Este es el método más riguroso y preferible. Consiste en dividir la escala en dos partes equivalentes y hallar la correlación de los resultados obtenidos en las dos aplicaciones. Presenta la ventaja de que pueden ser aplicadas las dos mitades al grupo en una sola sesión, con lo que se obvian los inconvenientes de las variaciones y las influencias distintas de una aplicación a otra.

Este método exige, claro es, que las dos mitades sean equivalentes.

El procedimiento más recomendable es separar en la escala los items pares de los items impares después de haber sido aplicada la escala en su totalidad, *como un todo.*

De esta forma se tendrán dos series de puntuaciones referentes a los items pares e impares. La correlación entre ellas es el coeficiente de seguridad de la escala, que resulta de la fórmula que sigue:

$$S = \frac{2\rho}{1+\rho}$$

ρ (rho) es el coeficiente de correlación obtenido.

Además de los procedimientos expuestos existen fórmulas para hallar la seguridad de una escala que se basan simplemente en el análisis estadístico de los mismos items de la escala cuya seguridad se busca.

Entre ellas es conocida la llamada fórmula 20 de Kuder-Richardson, modificada después por Horst.

Más simple y fácil es la llamada técnica de Roulon (Vid. Gonzalvo, 1978, p. 134), cuya fórmula es la siguiente:

$$r_{tt} = 1 - \frac{\sigma^2_{dif.}}{\sigma^2_{total}}$$

donde σ^2_{dif} = varianza de las diferencias entre las puntuaciones pares e impares, y σ^2_{total} = varianza de ambas mitades. (Ver ejercicio 146)

Fórmula de Spearmann Brown

La seguridad de una escala depende del número de sus items.

Cuanto mayor sea este número, mayor será su seguridad. Existe una fórmula llamada de Spearmann Brown que permite calcular la seguridad que se puede alcanzar mediante un aumento determinado de items homogéneos de la escala. Es la siguiente:

$$S_n = \frac{n \cdot r_{11}}{1 + (n-1) \cdot r_{11}}$$

r_{11} = seguridad de la primera escala.
n = número de veces de aumento.

Despejando n en esta fórmula, se obtiene esta otra que nos indica el número de items que hay que aumentar a una escala para obtener una seguridad dada o deseada:

$$n = \frac{S_n\ (1 - r_{11})}{r_{11} \cdot (1 - S_n)}$$

(Ver ejercicio 142)

VALIDEZ DE UNA ESCALA

Según el profesor Bugeda (1974, págs. 191 y ss.), la validez de una escala indica el grado de acuerdo de una escala con criterio que es exterior a ella.

Ello quiere decir que si, por ejemplo, hemos construído una escala para medir el autoritarismo y queremos determinar su validez, es necesario que tengamos un criterio de contraste, una medida previa del autoritarismo del grupo o la consigamos por algún procedimiento ya conocido.

El coeficiente de validez de una escala es el que resulta de correlacionar los resultados obtenidos con su aplicación con los del criterio seleccionado.

Este procedimiento no es muy práctico en cuanto requiere conocer la validez del criterio de comparación, que generalmente no es conocido ni fácil de conocer.

En su lugar se suelen aplicar otros procedimientos menos rigurosos entre los que cabe destacar los "know groups" y la "construct validity"

En el primer caso, la escala se aplica como prueba a grupos conocidos que se sabe poseen unos valores muy altos y muy bajos respecto a la actitud investigada. Igualmente se aplica un grupo normal. Si de los resultados obtenidos se deriva que los valores medios de los dos grupos difieren significativamente, es un indicio de validez de la escala.

La validez construida se establece comprobando el grado de relación entre las predicciones derivadas de la definición teórica de las variables medidas por el instrumento de observación con los resultados obtenidos mediante la aplicación de éste.

De acuerdo con Mannheim y Rich (1986, 87) esta validación implica el razonamiento siguiente: "Si el concepto X tiene una relación positiva con el concepto Y, una relación negativa con el concepto Z (como dice nuestra teoría) también será cierto que las calificaciones en una medida válida de X tendrán una relación positiva con las calificaciones en una medida válida de Y y una relación negativa con las calificaciones en una medida válida de Z".

También se pueden validar las escalas, aunque es un procedimiento penoso de aplicar, mediante el análisis de si la conducta de aquellos a quienes se han aplicado, concuerda con los resultados obtenidos en las escalas. A este procedimiento se le llama validez predictiva.

La validez de una escala está también relacionada con su seguridad. Por tanto, a mayor seguridad, también mayor validez. Así como aumentando el número de item se puede aumentar la seguridad de una escala, igualmente aumentando el número de preguntas o proposiciones

homogéneas se puede aumentar también la validez. La fórmula es la siguiente:

$$V_{xy} = \frac{v_{xy}}{\sqrt{\frac{1 - r_{11}''}{n} + r_{11}''}}$$

V_{xy} = Validez buscada.
v_{xy} = Validez inicial.

(Ver ejercicio 144)

LAS CURVAS DE OPINION PUBLICA

En relación al análisis de los resultados de las escalas de actitudes presentan interés especial las llamadas curvas de opinión pública.

Como escribe el profesor Bugeda (1974, p. 208), imaginemos, por ejemplo, que una encuesta o escala de opinión sobre los esquimales se aplica a un grupo de universitarios españoles.

Lo más normal es que muy pocos de ellos, si algunos, manifiesten una máxima aceptación de los esquimales. También será lo más probable que muy pocos les odien de tal forma que expresen en la encuesta su rechazo total hacia ellos.

Los grados de aceptación se acumularán más bien en las categorías intermedias. Girarán en torno a la indiferencia y una moderada simpatía o antipatía hacia los esquimales.

La representación histográfica de una opinión semejante daría lugar a la formación de una curva normal de probabilidades, o sea de las llamadas de campana de Gauss. En realidad ésta es la representación típica de cualquier fenómeno natural.

Pero imaginemos, por el contrario, que a ese mismo grupo de jóvenes españoles le planteamos, en vez del indiferente tema de los esquimales, una cuestión que les apasione, como el feminismo, el divorcio, el aborto, la selectividad universitaria, etc.

Lo que ocurriría ahora sería una cosa muy diferente.

Las categorías de indiferencia estarán probablemente vacías. Tampoco estarán muy llenas las que signifiquen aceptación o rechazo moderados. Lo más probable es que las opiniones se concentren en las categorías extremas, máxima aceptación o máximo rechazo o en ambas.

Esto indica que el grupo está dividido en cuanto a opinión o actitud sobre el tema, aunque desde luego, esta división no tiene por qué ser en dos partes iguales, ya que normalmente uno de los grupos de actitud predominará en cada tema sobre el otro.

Esta partición polémica de un grupo social en torno a una cuestión que considera fundamental, es precisamente la señal de que estamos ante la opinión pública.

De ello se deduce que la opinión privada y la opinión pública dan lugar a distintas clases de curvas.

Según he indicado, la curva típica de la opinión privada es la curva de Gauss, llamada también normal. Las curvas típicas de la opinión pública no son del tipo de campana, como las de las opiniones privadas, sino que aparecen más bien bajo la forma de campana invertida, o de una rama de ésta. Por ello algunos las llaman curvas en U, J o ʃ.

Las escalas de actitudes se refieren normalmente a opciones fundamentales de los individuos, de ahí que interese examinar, en relación a los resultados de su aplicación, si existe o no opinión pública en el grupo observado.

A este efecto, se pueden formar las curvas del grupo para la escala en su conjunto o respecto a un item determinado, así como las curvas de frecuencias acumulativas.

Con relación a estas curvas de opinión pública se plantea también el problema estadístico de determinar si la anormalidad de distribución que se interpreta como de opinión pública es realmente significativa y no se debe simplemente al azar.

Para ello se puede utilizar uno de los tests estadísticos más conocidos y utilizados en las ciencias sociales: el ji cuadrado.

Su fórmula es:

$$\chi^2 = \Sigma \frac{(e_t - t_n)^2}{t_n}$$

en la que e_t = a las medidas efectivas obtenidas.
t_n = las medidas teóricas o normales.

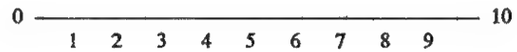
Para aplicar este tests a efectos de contrastar si la distribución obtenida efectivamente en una escala difiere o no significativamente de la distribución normal de probabilidad, es preciso hallar primero las frecuencias teóricas o normales, para lo cual se utiliza la tabla de áreas de la curva normal (Vid. ejercicios núms 137 a 140)

EJERCICIOS

Para un cuestionario sobre cuestiones familiares, redactar un ítem que constituya una escala estimativa, representada por una línea continua dividida de 0 a 10 puntos para que el encuestado sitúe su posición favorable o desfavorable a la admisión del aborto.

RESPUESTA

"Suponiendo que la actitud hacia el aborto está representada por una escala formada por una línea continua dividida en diez partes iguales, en la que el cero corresponde a una actitud completamente contraria al aborto, y el diez a una actitud plenamente favorable ¿en qué punto de la escala situaría usted su posición respecto a esta cuestión?



132

Dentro de una investigación de sociología religiosa, que se está efectuando, se ha proyectado utilizar una escala arbitraria de puntos para medir la actitud progresista-integrista en el campo religioso. Enunciar una serie de palabras de tipo integrista o progresista que puedan despertar una repulsa en las personas de tendencia contraria y servir como punto de partida para la construcción, previas las pruebas oportunas, para comprobar su validez y seguridad, de la escala de que se trata, redactando asimismo la fórmula de presentación de la escala e indicando el modo de evaluar los resultados obtenidos con su aplicación a cada sujeto.

Para ello se tendrá en cuenta:

1. Que las palabras de signo integrista deberán ser tantas como las de signo progresista y respecto a cada punto se presentarán las palabras que representen las dos tendencias contrarias.

2. Se deben colocar en la escala las palabras elegidas al azar, pero de modo que no aparezcan juntas las dos contradictorias sobre un mismo punto.

RESPUESTA

Serie de palabras de la escala:

Contrarreforma; comunidades de base; cardenalato, ecumenismo; pueblo de Dios; sínodo de obispos; infalibilidad pontificia; sumisión a la jerarquía; ordenación de mujeres; Concilio de Trento; índice de libros prohibidos; curas obreros; oración comunitaria; designación de los obispos por los fieles; meditación; misa en latín; paraliturgia.

Presentación: Se podría utilizar la siguiente, que figura en la escala de Watson para medir la actitud político-moral. (Cfr. *Ob. cit.* del profesor Bugeda Sanchiz. Segunda edición. Madrid, 1974, pág. 135.)

"Lea las palabras catalogadas más abajo, examine cada una rápidamente; si le sugiere que es más desagradable que agradable, táchela; puede tachar muchas o pocas palabras, obre tan rápidamente como pueda, pero con la seguridad de que tacha aquella palabra que es más molesta que agradable, más angustiante que excitante, más desagradable que atractiva.

Puntuación: Para la puntuación se podría dar inicialmente a cada individuo a quien se aplique la escala tantos puntos como palabras o proposiciones comprenda aquélla, a la que se sumaría un punto por cada palabra tachada de signo integrista y se restaría otro por las de carácter progresista. La mayor o menor puntuación obtenida por encima o debajo del número de puntos inicial concedido indicaría la intensidad o el grado de integrismo del sujeto en cuestión.

133

Supuesto que se tiene que realizar una investigación sobre la aceptación del régimen democrático y del franquista, se pide proyectar una escala de clasificación que refleje la actitud hacia las instituciones del franquismo, teniendo en cuenta que:

1. Esta escala consiste en proponer una lista de instituciones del régimen mezcladas al azar con otras de regímenes opuestos o distintos.

2. Esta lista de instituciones hay que presentarlas de forma que su disposición no pueda influir en el orden que el individuo a quien se aplique la escala elija.

3. Es necesario que vaya acompañada, como toda escala, de una presentación, en la que se solicite la cooperación y se indique la manera de cumplimentar la escala.

4. Hay que idear un sistema de puntuación de las clasificaciones realizadas por cada sujeto a quien se aplique la escala.

5. El encuestado debe clasificar la lista ordenando según sus preferencias las instituciones propuestas.

RESPUESTA

SERIE DE INSTITUCIONES DE LA ESCALA

Cámara de Diputados, Consejo del Reino, Caudillo, Constitución democrática, Cortes Españolas, Sindicatos libres, Tribunal de Orden Público, Ministro de Relaciones Sindicales, Consejos Obreros, Consejo Nacional del Movimiento, jurados populares, partidos políticos.

Presentación: Pertenecemos al Instituto... y estamos realizando una investigación para conocer las preferencias políticas de los españoles. Por ello le rogamos ordene todas las instituciones que se enumeran por orden de preferencia, poniendo las que cree mejores o le agradan más para España en primer lugar, y las que considera peores o más rechazables las últimas, según le agradan o desagradan más o menos. Su información es rigurosamente secreta. No debe firmar ni poner su nombre en ningún caso.

Puntuación: Como se pretende medir la actitud política hacia las instituciones del franquismo, sólo se tendrán en cuenta éstas en la puntuación. A cada institución de este tipo se le podría asignar la puntuación correspondiente al número de orden que ocupe en la clasificación realizada por el sujeto. Sumados los puntos de las seis instituciones, nos daría la puntuación total. Esta se hallaría entre 21 puntos que corresponden a la colocación de todas las instituciones del Movimiento antes que las opuestas y expresa la máxima adhesión a ellas y 57 puntos resultarían si todas las instituciones indicadas se pusieran después que las opuestas.

134

En una investigación sociológica sobre la Universidad española, que se está realizando entre universitarios y posgraduados, se ha pensado aplicar un diferenciador semántico para conocer el concepto que los encuestados tienen formado de la universidad actual y su imagen ideal de la misma.

Preparar el proyecto de dicho diferenciador semántico, teniendo en cuenta:

1. Que, como es sabido, esta técnica escalar consiste en proponer una serie de adjetivos en forma bipolar, es decir, comprendiendo el positivo y su contrario, por ejemplo, fuerte-débil.

2. Cada uno de los polos de la forma adjetival en cuestión se coloca en los extremos de una escala formada por una línea continua, dividida en una serie de grados o rangos. Por ejemplo, en la escala siguiente el grado extremo de abstracta sería muy abstracta, y los siguientes, bastante abstracta, algo abstracta, ni abstracta ni concreta, algo concreta, bastante concreta y muy concreta.

3. A la persona que se le aplica se le pide que señale el lugar que dentro de la escala y con relación a los adjetivos correspondientes ocupa la persona, institución o cosa concreta investigada.

4. A la vez se le puede pedir, especialmente en el caso de instituciones, que indique el lugar ideal que en la misma escala y respecto a los mismos objetivos debería ocupar, en su opinión, tal institución.

RESPUESTA

En el caso propuesto es preciso buscar una serie de adjetivos positivos y negativos que puedan calificar la universidad y disponerlos en la forma indicada.

Se inserta a continuación un conjunto de estos adjetivos, en forma de diferenciador semántico, basado con algunas modificaciones en el preparado para un trabajo similar de J.M. Favergue inserto en la *Revue de l'Institut de Sociologie*, 1971, 4:

Presentación: Pertenecemos al Instituto... y estamos realizando una investigación para conocer la imagen actual e ideal de la Universidad española. Por ello le rogamos que en el adjunto diferenciador semántico marque con una A en la casilla que cree corresponde a la Universidad actual y con una I la que a su juicio correspondería a

Abstracta								Concreta
Controlada por el Estado								Independiente
Individualista								Comunitaria
Apolítica								Políticamente activa
Sabia								Crítica
Vinculado al capital								Vinculada a la clase obrera
Tecnocrática								Ideológica
Cuidadosa de la tradición								Cuidadosa del cambio
Profesional								Cultural
Al margen de la sociedad								Motor de la sociedad
Científica								Educativa
Técnica								Humanista

la Universidad ideal. Para ello tenga en cuenta que los valores de la escala se entienden ordenados de más a menos en sentido de izquierda a derecha para los adjetivos de la izquierda y de más a menos en el sentido de derecha a izquierda para los de la derecha.

Puntuación: La unión mediante líneas continuas y trazos para la Universidad actual y para la Universidad ideal, respectivamente, de las casillas elegidas respecto a cada adjetivo proporciona una representación gráfica de la imagen sustentada por cada sujeto respecto a la Universidad ideal y la actual. La puntuación se podría realizar asignando puntos a cada casilla del uno al siete, de izquierda a derecha. Esta puntuación representaría la actitud respecto a la dimensión a que se refiere cada par de adjetivos. Para obtener la puntuación total de la escala para cada individuo se sumarían todos los puntos correspondientes a las casillas señaladas tanto respecto a la Universidad actual como respecto a la Universidad ideal. Dado que a la izquierda figuran los adjetivos que se pueden considerar más conservadores y a la derecha los más progresistas, una puntuación total baja respecto a la Universidad actual significaría una imagen conservadora de ésta, así como una puntuación total alta en la Universidad ideal, una imagen ideal progresista de ésta.

135

Redactar un escalograma de seis items sobre la práctica religiosa de los católicos.

RESPUESTA

¿Va usted a misa y comulga todos los días?

Sí.
No.

¿Comulga usted por lo menos una vez a la semana?

Sí.
No.

¿Confiesa y comulga por lo menos una vez al mes?

Sí.
No.

¿Oye misa todos los domingos y días de precepto, salvo que se lo impida algún motivo justificado?

Sí.
No.

¿Asiste a misa únicamente alguna vez durante el año?

Sí.
No.

¿Ha bautizado a sus hijos o los bautizaría si no los tiene y los tuviera?

Sí.
No.

136

Redactar otro escalograma sobre medidas a adoptar contra los neonazis.

RESPUESTA

Los neonazis deben ser condenados a muerte.

De acuerdo.
En desacuerdo.

Deben ser encarcelados mientras no renieguen de su ideología.

De acuerdo.
En desacuerdo.

Deben ser arrojados fuera del país.

De acuerdo.
En desacuerdo.

No se les debe permitir trabajar en ninguna empresa pública ni privada.

De acuerdo.
En desacuerdo.

No se les debe reconocer sus derechos políticos.

De acuerdo.
En desacuerdo.

No se les debe permitir ocupar cargos ni funciones públicos.

De acuerdo.
En desacuerdo.

137

Se ha aplicado la escala radicalismo-conservadurismo de Netler, de tipo Likert, a un grupo de estudiantes de una clase de veinte alumnos.

Las puntuaciones medias de los alumnos del grupo, agrupadas por los intervalos de frecuencias que se indican, obtenidas de dividir la suma total de puntos de la escala por 14, número de ítems de la escala, fueron las siguientes:

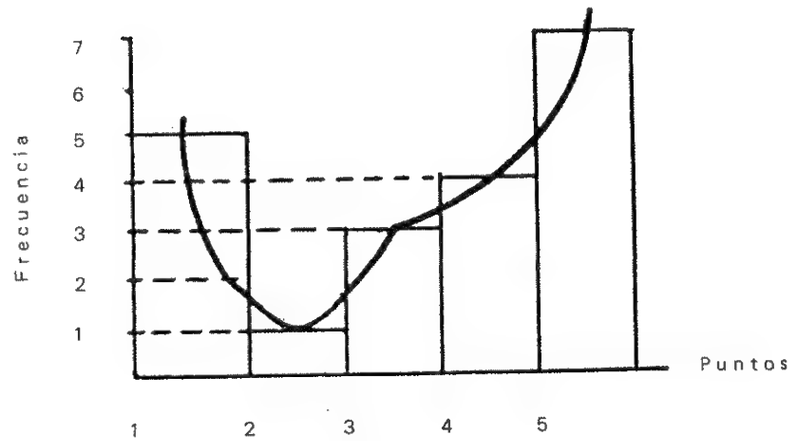
Número de alumnos	*Puntos*
5	0 a 1
1	1 a 2
3	2 a 3
4	3 a 4
7	4 a 5

Se pide analizar los resultados de esta aplicación de la escala, construyendo el diagrama que exprese la posición del grupo respecto a la actitud medida por la escala e interpretarlos.

RESPUESTA

Basta llevar a un eje de coordenadas los valores indicados, consignando los intervalos de clase de las puntuaciones en el eje de las X y las frecuencias en el de las Y. Se obtendrá así un diagrama de barras. Para diseñar la curva se unen los puntos medios del lado opuesto a la base de las barras formadas.

Interpretación: La curva obtenida tiene forma de U, clásica de las curvas de opinión pública, e indica que la actitud del grupo está dividida en cuanto predominan las posiciones extremas.



138

Dadas las frecuencias de las puntuaciones medias obtenidas señaladas en el ejercicio anterior, se pide hallar las frecuencias teóricas que corresponderían a las primeras si su distribución se ajustara a la curva normal o de Gauss, en el caso de una desviación —supuestas— de 2,50 y de 2, respectivamente, teniendo en cuenta que para ello es preciso:

1. Hallar las marcas de clase.

2. Dividir la diferencia entre la marca de clase y la media dada por la desviación típica.

3. Hallar los valores que en la tabla de áreas de la curva normal corresponden a los cocientes de la división anterior, buscando éstos en la columna x/σ de dicha tabla. El área de la curva normal correspondiente se encuentra en la columna 0 si el cociente de x/σ sólo tiene una cifra decimal, por ejemplo, si es 1,0 ó 1,50; pero si es 1,55 figura en la intersección de la fila correspondiente a 1,50 y la columna 5, etc.

4. Hallar el área de la curva normal que corresponde a cada intervalo de clases, restando cada valor obtenido en la columna de las áreas de la columna normal del que le precede.

5. Obtener las puntuaciones teóricas multiplicando el área de la curva normal que corresponde a cada intervalo por N, o el número de individuos del grupo, en este caso 20.

RESPUESTA

Disposición de los cálculos:

Intervalos de clase	*Marcas de clase, x*	$\frac{x - \bar{x}}{\sigma}$	*Area curva normal*	*Area de cada intervalo*	*Frecuencia teórica Col. 5 × 20*	*Frecuencia empírica*
0 — 1	0,50	1	0,3413	—	—	—
1 — 2	1,50	0,5	0,1915	0,1498	2,9	7
2 — 3	2,5	0	0	0,1915	3,8	1
3 — 4	3,5	0,5	0,1915	0,1915	3,8	3
4 — 5	4,5	1	0,3413	0,1498	2,9	4
5 — 6	5,5	1,5	0.4332	0,0919	1,8	7

Interpretación: Se ve, como era de esperar, dada la curva hallada en el ejercicio anterior, que ambas distribuciones, la teórica y la empírica, son muy diversas. Esta divergencia indica que en el grupo existe opinión pública en relación a la actitud que mide la escala en cuestión.

139

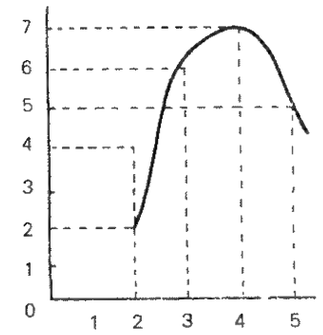
Las puntuaciones dadas por el grupo en cuestión al ítem de la escala del ejercicio anterior siguiente: "Los beneficios de las grandes empresas deben estar absolutamente controlados por el Gobierno." Se ha distribuido de la siguiente forma:

Número de alumnos	*Puntos*
5	5
7	4
6	3
2	2

Se pide construir la curva que exprese la actitud del grupo en relación a dicho ítem e interpretarla.

RESPUESTA

El procedimiento es igual que en el ejercicio anterior, con la diferencia que en aquél, al referirse a las puntuaciones medias de cada sujeto en la escala aparecían acumuladas las frecuencias según intervalos de clase.



La curva resultante es del tipo de las de campana, o de Gauss. Indica que no existe en el grupo una opinión definida sobre este ítem, por predominar en él las posiciones centrales. Hay que tener en cuenta el desplazamiento de la curva a la derecha.

140

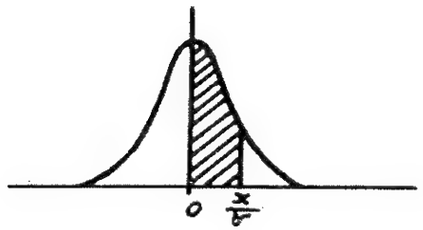
Dada la distribución de frecuencias de las puntuaciones del ítem de la escala a que se refiere el ejercicio anterior, se pide hallar las frecuencias teóricas que corresponderían si la distribución empírica se ajustara a la curva normal, siendo la media 3 y la división típica 2 –valores supuestos–. Interpretar los resultados.

RESPUESTA

Disposición de los cálculos:

Puntuación	$\frac{x - \bar{x}}{\sigma}$	*Area curva normal*	*Area de cada intervalo*	*Frecuencia teórica: Col. 4 × 20*	*Frecuencia empírica*
5	1	0,3413	—	—	—
4	0,50	0,1915	0,1498	2,90	5
3	0	0	0,1915	3,80	7
2	0,5	0,1915	0,1915	3,8	6
1	1	0,3413	0,1408	2,9	2

Interpretación: De los datos obtenidos se deriva una cierta similitud entre ambas distribuciones de frecuencias, en armonía con la curva de frecuencias empíricas. Para ver si se trata o no en realidad de dos distribuciones diferentes estadísticamente, es preciso realizar la prueba del χ^2 a que se refieren los ejercicios 272, 273 y 274. Si las dos distribuciones no son diferentes según esta prueba, indica la no existencia de opinión pública en el grupo en cuestión en relación a este ítem de la escala.



χ/σ	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0159	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2930	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3718	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4083	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4430	0,4441
1,6	0,4452	0,4463	0,4474	0,4485	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4758	0,4762	0,4767
2,0	0,4773	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4865	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4980	0,4980	0,4981
2,9	0,4981	0,4982	0,4983	0,4984	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,49865	0,4987	0,4987	0,4988	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990
3,1	0,49903	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993

χ/σ	0	1	2	3	4	5	6	7	8	9
3,2	0,4993129									
3,3	0,4995166									
3,4	0,4996631									
3,5	0,4997674									
3,6	0,4998409									
3,7	0,4998922									
3,8	0,4999277									
3,9	0,4999519									
4,0	0,4999683									
4,5	0,4999966									
5,0	0,4999997133									

Esta tabla indica la proporción del área bajo la curva "normal" comprendida entre la media y una ordenada levantada a una cierta distancia de la media misma siendo tal distancia expresada en iguales unidades que la desviación típica (σ).

Es de notar que se toma sólo en consideración un lado de la curva, esto es, 0,5 del área de la curva (1 = área total). Por ejemplo, entre la media y 2 σ (o bien entre la media y $-$ 2 σ) está situado el 47,725 por 100 del área de la curva normal. Por consiguiente, entre la media y $\pm$ 2 σ está situado el 95,45 (47,725 x 2) por 100 de dicha curva.

Modo de empleo. Si la duración media de los radios de 1.000 familias tomadas al azar es de 5,7 años, y si σ = 2,8 años, podemos calcular el área de la curva normal comprendida, por ejemplo entre 4 y 6 años. En este caso tendremos:

$$\frac{6{,}0 - 5{,}7}{\sigma} = 0{,}11$$

de σ sobre la media:

$$\frac{4{,}0 - 5{,}7}{\sigma} = 0{,}61 \text{ de } \sigma \text{ sobre la media.}$$

La tabla nos muestra que el 4,38 por 100 del área bajo la curva, está situada entre la media y 0,11 de σ y que 22,9 por 100 del área bajo la curva, está situada entre la media y 0,61 de σ. Sumando los dos porcentajes anteriores se deduce que el 27,29 por 100 de las radios de la muestra están comprendidas entre 4 y 6 años. Prácticamente los valores de la tabla se usan también para distribuciones que no son normales, salvo casos de distribuciones completamente diferentes.

Nota.– 0,11 de σ se encuentra partiendo de 0,1 en la primera columna y recorriendo horizontalmente hasta encontrar la columna encabezada 1 (0,0438).

141

Se está construyendo una escala de actitudes de tipo Thurstone para medir la actitud progresismo-conservadurismo y entre los items de la misma sometidos, con el

objeto de probarla, a un grupo de 50 jueces se encuentra el siguiente: "La jornada laboral fuera del hogar de la mujer casada debería ser inferior a la normal".

La evaluación objetiva dada por el grupo de jueces, siendo 1 el máximo progresismo y 11 el máximo conservadurismo, al ítem fue la siguiente:

Número de encuestados	*Frecuencias acumuladas*	*Puntos dados al ítem*
5	5	3
10	15	5
15	30	6
9	39	8
4	43	9
5	48	10
2	50	11

Valor escalar del ítem, 5,5.

Determinar el valor escalar del ítem y su dispersión, utilizando al efecto la mediana y la dispersión intercuártica en lugar de la $\bar{x}$ y la σ, construyendo al efecto la curva de frecuencias acumuladas.

Decidir, finalmente, si dada la dispersión del ítem que resulte según el coeficiente de ambigüedad, el ítem es o no aceptable para medir la actitud en cuestión.

RESPUESTA

Construcción de la curva de frecuencias acumuladas:

Tanto por ciento que corresponde a cada individuo del grupo: 100 : 50 = 2

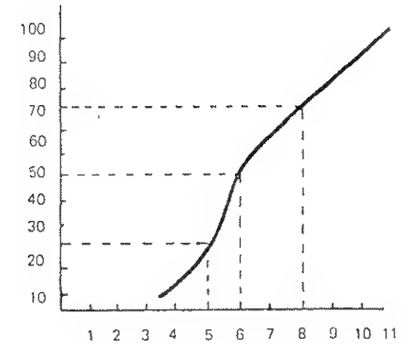
Tanto por ciento que corresponde a cada grado de la lista de frecuencias acumuladas: 5 x 2 = 10; 15 x 2 = 30; 30 x 2 = 60; 78, 86, 96, 100.

Llevando estos tantos por ciento acumulados al eje de las Y y los valores al de la X, se tiene la curva siguiente:

El valor escalar del ítem, que es la puntuación o valor que corresponde en el eje de las X a las frecuencias acumuladas del 50%, alcanza en este caso el valor aproximado de 5.7.

En cuanto al coeficiente de ambigüedad, dado que la puntuación o valor en el eje X, que corresponde a las frecuencias acumuladas de 75 y 25%, es de aproximadamente de 7,6 y 4,5, es 3,31 o la diferencia entre ambas puntuaciones.

La dispersión resultante es superior a 2 luego se trata de un ítem que no representa adecuadamente a la actitud medida, ya que el acuerdo de los jueces en relación a su significación respecto a dicha actitud no es todo lo acusada que debiera. Luego se debe rechazar el ítem por no ser lo suficientemente idoneo para formar parte de la escala.



142

Se tiene una escala de 20 items con una seguridad calculada de 0,60, ¿qué seguridad tendrá una escala tres veces mayor de items homogéneos?

RESPUESTA

La fórmula a aplicar es la siguiente, llamada de Spearman-Brown:

$$S_n = \frac{n \cdot r_{11}}{1 + (n - 1) \cdot r_{11}}$$

en la que r_{11} es la seguridad de la primera escala y n el número de veces que se aumenta la escala.

Aplicando la fórmula se tiene:

$$S_n = \frac{3 \times 0{,}60}{1 + (2 \times 0{,}60)} = \frac{1{,}80}{2{,}20} = 0.81$$

Si se triplica, pues, el número de items, la seguridad pasará de 0,60 a 0,81.

143

Una escala de 25 ítems tiene un coeficiente de seguridad de 0,50, cuántos ítems habrá que aumentar para lograr un coeficiente de seguridad de 0,70.

RESPUESTA

Fórmula a aplicar:

$$n = \frac{S_n \cdot (1 - r_{11})}{r_{11} \cdot (1 - S_n)}$$

Aplicando la fórmula se tiene:

$$n = \frac{0{,}70 \times (1 - 0{,}50)}{0{,}50 \times (1 - 0{,}70)} = \frac{0{,}35}{0{,}15} = 2{,}33$$

La nueva escala debe ser 2,33 veces mayor que la inicial, o sea, ha de tener 2,33 x 25 = 58 ítems, y como ya tenía 25, hay que añadirla 33 más.

144

Una escala formada por 15 ítems y con un coeficiente de seguridad de 0,30 tiene una validez de 0,35. Indicar: 1. cuál será la validez de una escala paralela de 30 ítems, y 2. hallar la validez máxima a que se podría llegar si se sigue aumentando el número de sus ítems.

RESPUESTA

Téngase en cuenta que la fórmula para la primera cuestión es:

$$V_{xy} = \frac{v_{xy}}{\sqrt{\frac{1 - r_{11}}{n} + r_{11}}}$$

V_{xy} es la validez buscada.
v_{xy} es la validez de la escala inicial.
r_{11} es la regularidad de la escala inicial.
n es la relación entre la segunda escala y la primera.

La fórmula a aplicar en la segunda cuestión es:

$$\text{Validez máxima } V_{xoy} = \frac{v_{xy}}{\sqrt{r_{11}}}$$

Sustituyendo en la primera fórmula se tiene:

$$V_{xy} = \frac{0{,}35}{\sqrt{\frac{1 - 0{,}30}{2} + 0{,}30}} = \frac{0{,}35}{0{,}80} = 0{,}43$$

La comparación de la validez de la primera escala con la de la segunda muestra un aumento de la validez en una proporción muy inferior al aumento del número de ítems.

Sustituyendo igualmente en la segunda fórmula se tiene:

$$V_{xoy} = \frac{0{,}35}{\sqrt{0{,}30}} = \frac{0{,}35}{0{,}54} = 0{,}64$$

Por tanto, sea cual sea el número de ítems que añadamos a la escala inicial, la validez máxima que se puede conseguir en este caso es 0,64.

145

Se esta construyendo una escala de actitudes de tipo Likert y con el objeto de comprobar la consistencia de los ítems ideados para la misma se ha aplicado la escala a un grupo de diez sujetos. Los resultados obtenidos en la aplicación total de la escala y en relación al ítem 5 son los siguientes:

Sujetos	*Puntuación total*	*Puntuación en el ítem 5*
A	45	5
B	42	5
C	35	4
D	35	4
E	20	1
F	39	4
G	33	3
H	40	4
I	21	1
J	25	2

Determinar el grado de consistencia de la puntuación dada al ítem en relación a la puntuación total de la escala.

RESPUESTA

Como se ha indicado al tratar de la construcción de las escalas Likert, la consistencia de los ítems en este tipo de escalas se determina correlacionando la puntuación total de cada sujeto menos la del ítem en cuestión, en este caso el 5, con la puntuación en este mismo ítem.

El coeficiente de correlación pertinente, al tratarse en las escalas de actitudes de Likert de escalas de medida de nivel ordinal, es el rho de Spearman. (Sobre este coeficiente ver ejercicios 264 y 268).

La fórmula es la siguiente: $\rho = 1 - \dfrac{6\,(\Sigma d)^2}{N\,(N^2 - 1)}$

Punt. total	*Su orden en rangos*	*Punt. ítem 5*	*Su orden en rangos*	*Diferencia rangos*	*Dif.*^{2}
1	2	3	4	(2–4)	$(2-4)^2$
45	1	5	1.5	.5	.25
42	2	5	1.5	.5	.25
35	5.5	4	4.5	1.	1.
35	5.5	4	4.5	1.	1.
20	10.	1	9.5	.5	.25
39	4	4	4.5	.5	.25
33	7	3	7	0	0.
40	3	4	4.5	.5	.25
22	9	1	9.5	.5	.25
27	8	2	8	0	0

$$\rho = 1 - \frac{6\,(5)}{10\,(100\text{-}1)} = 1 - 3\text{0}/990 = .97 \quad \Sigma\,(\text{dif})^2\ 5.$$

Como se ve el coeficiente obtenido es muy alto. La prueba realizada muestra que la consistencia interna de este ítem en relación a la escala total, es muy elevada.

146-8

Dados los siguientes resultados obtenidos por la aplicación de una escala de actitudes a un grupo apara hallar su seguridad,

Número del ítem	*Suma de puntos dados al ítem por el grupo*
1	8
2	9
3	7
4	6
5	3
6	4
7	3
8	3

Hallar la seguridad de la escala aplicando la técnica de Roulon.

RESPUESTA

La fórmula a aplicas es $r_{tt} = 1 - \dfrac{\sigma^2_{dif}}{\sigma^2_{tot}}$

donde σ^2_{dif} es la varianza de las diferencias entre las puntuaciones de los ítems pares e impares y σ^2_{tot} la varianza de las puntuaciones de la escala completa.

Para hallar esta varianza podemos formar la siguientes tabla de datos.

Items	*Puntos escala total, X*	*X*	*Puntos ítems pares*	*Puntos ítems impares*	*Dif. pares impares*	*Dif.*^{2}
1	8	64	8			
2	9	81		9	–1	1
3	7	49	7			
4	6	36		6	1	1
5	6	36	6			
6	4	16		4	2	4
7	3	9	3			
9	3	9		3	0	0
	46	300			2	6

$$\sigma^2_{tot} = \frac{300}{8} - \left(\frac{46}{8}\right)^2 = 37.5 - 33.06 = 4.43$$

$$\sigma^2_{dif} = \frac{6}{4}\left(\frac{2}{4}\right)^2 = 1.50 - .25 = 1.25$$

$$r_{tt} = 1 - \frac{1.25}{4.43} = 1 - .28 = .72$$

Seguridad de la escala, $r_{tt} = .72$.

Sección 5

CLASIFICACION

20. CODIFICACION Y TABULACION

NOCION Y FASES DE LA CLASIFICACION

Terminada la etapa de la investigación de recogida de datos, la aplicación a la población objeto de la investigación de los instrumentos de observación se concreta en una masa informe de datos individuales, sin agrupación alguna y, por tanto, carentes en tal estado de significación sociológica, en cuanto el objeto de ésta son los fenómenos colectivos y no los particulares.

Por ello, la etapa esencial de la investigación social, siguiente a la observación, es la clasificación, o sea, la agrupación de los datos recogidos referentes a cada variable objeto de estudio y su presentación conjunta, bien separadamente o relacionados con los de otras variables.

El objeto de la clasificación es, pues, reflejar, previa su diferenciación, la dimensión colectiva de los datos recogidos en la observación y con ello poner de manifiesto las uniformidades, semejanzas y diferencias de los fenómenos sociales.

Mediante la clasificación se pasa del dato en bruto al dato procesado.

Las secuencias de este proceso y sus operaciones aparecen en el siguiente esquema de Galtung (1971, p. 209):

Dato en bruto → Puntaje → Puntaje sobre hoja de codificación → Perforación → Dato procesado

La clasificación comprende, pues, las siguientes operaciones:

1. Codificación, en la que se dan puntajes a las respuestas en conformidad a las instrucciones de codificación, por ej. al margen del cuestionario.

2. Transferencia, en la que se transfieren los puntajes a hojas de codificación, donde los puntajes se presentan de forma que facilite su perforación.

3. Perforación, en la que los datos se transfieren desde las hojas de codificación a fichas que se utilizan para realizar el trabajo mecánico de refundir los datos.

Sección 5

CLASIFICACION

20. CODIFICACION Y TABULACION

NOCION Y FASES DE LA CLASIFICACION

Terminada la etapa de la investigación de recogida de datos, la aplicación a la población objeto de la investigación de los instrumentos de observación se concreta en una masa informe de datos individuales, sin agrupación alguna y, por tanto, carentes en tal estado de significación sociológica, en cuanto el objeto de ésta son los fenómenos colectivos y no los particulares.

Por ello, la etapa esencial de la investigación social, siguiente a la observación, es la clasificación, o sea, la agrupación de los datos recogidos referentes a cada variable objeto de estudio y su presentación conjunta, bien separadamente o relacionados con los de otras variables.

El objeto de la clasificación es, pues, reflejar, previa su diferenciación, la dimensión colectiva de los datos recogidos en la observación y con ello poner de manifiesto las uniformidades, semejanzas y diferencias de los fenómenos sociales.

Mediante la clasificación se pasa del dato en bruto al dato procesado.

Las secuencias de este proceso y sus operaciones aparecen en el siguiente esquema de Galtung (1971, p. 209):

Dato en bruto → Puntaje → Puntaje sobre hoja de codificación → Perforación → Dato procesado

La clasificación comprende, pues, las siguientes operaciones:

1. Codificación, en la que se dan puntajes a las respuestas en conformidad a las instrucciones de codificación, por ej. al margen del cuestionario.

2. Transferencia, en la que se transfieren los puntajes a hojas de codificación, donde los puntajes se presentan de forma que facilite su perforación.

3. Perforación, en la que los datos se transfieren desde las hojas de codificación a fichas que se utilizan para realizar el trabajo mecánico de refundir los datos.

4. Tabulación, en la que se hace reaparecer los datos bajo una forma u otra a través de las máquinas o manualmente.

Antes de empezar la codificación es necesario realizar otra operación previa que los ingleses llaman "editing". Esta operación consiste simplemente en la revisión detenida de los cuestionarios o cuadros para detectar y eliminar en lo posible los errores y omisiones que pueden presentar.

En síntesis, en esta revisión, como dice M. Parten (1958, p. 453) en relación a los cuestionarios, debe comprobarse que los datos que han de ser tabulados sean:

1. Lo más exactos y fidedignos posible.
2. Compatibles con los demás informes emitidos.
3. Totalmente uniformes en su anotación.
4. Completos.
5. Apropiados para la tabulación.
6. Ordenados de forma que sean fáciles de poner en clave y tabular.

NOCION DE CODIFICACION

La clasificación comprende dos operaciones fundamentales: la codificación y la tabulación. Las otras dos operaciones indicadas, la transferencia y la perforación, se pueden considerar accesorias de aquéllas.

La codificación es previa a la tabulación pero tiene su razón de ser en ésta. Para que pueda verificarse la agrupación de datos en que consiste la tabulación, es imprescindible identificar antes de manera inequívoca cada categoría de los datos contenidos en los documentos cumplimentados de observación.

El objeto de la codificación es hacer representar o traducir cada respuesta de los cuestionarios y cada dimensión de los cuadros de observación por códigos o indicaciones numéricas que faciliten la tabulación.

La codificación consiste en asignar, a todas y cada una de las categorías que comprende el cuestionario o el documento de observación, los números y signos correlativos que sean precisos, según el procedimiento de tabulación empleado, para hacer posible la agrupación de los datos de cada documentos con los similares de todos los demás.

La codificación permite concretar en números y signos todos y cada uno de los datos del cuestionario. Estos números se pasan después a una ficha mediante taladros o perforaciones de la misma.

Después de esta operación resultará que tendremos resumidos en unos cuantos taladros o perforaciones de una ficha todos los datos contenidos en cada cuestionario o cuadro cumplimentados.

La codificación, según se ve, es una labor instrumental, pero no por ello deja de tener importancia y de constituir una operación delicada. De su perfecta realización depende que en la tabulación no haya errores ni resulte falseada o viciada al atribuir los datos a grupos que no les corresponde.

Además, cuando se trata de preguntas de cuestionario abiertas o de hechos y datos recogidos sin formar categorías previas, la codificación requiere que todas las múltiples respuestas recogidas se clasifiquen en categorías fijadas previamente, operación que comporta la interpretación de las respuestas de acuerdo con el objeto de la investigación.

Aún cuando la codificación se puede aplicar a cualquier documento de recogida de datos, como el más común entre éstos es el cuestionario, es a él al que se hace principalmente referencia al tratar de ella.

La codificación de un cuestionario es distinta según los datos hayan de ser pasados a fichas para su tabulación manual o para su tabulación electrónica.

En el primer caso, las fichas tienen únicamente en el modelo más corriente 126 taladros, numerados correlativamente en los márgenes a lo largo de todo su contorno, con los que se hacen corresponder los números también correlativos asignados a las respuestas del cuestionario. El borde de los taladros, cuyo número coincide con el de la respuesta elegida, se corta o perfora.

La codificación electrónica es distinta según se utilicen fichas perforadas, cintas o discos magnéticos.

En este sistema de tabulación, la codificación suele suponer una doble numeración. Primera la correspondiente a la columna y segunda la referente a los espacios o dígitos pertinentes dentro de cada columna.

La primera se atribuye a las preguntas y la segunda a sus respuestas. (Ver ejercicio 149)

CATEGORIZACION DE PREGUNTAS ABIERTAS

Como orientación para verificar la categorización de las preguntas abiertas se pueden tener en cuenta las siguientes normas:

1. La base fundamental de esta operación debe ser el objetivo de la investigación y el material concreto que forman las respuestas.

2. Para conjuntar estas exigencias básicas es conveniente formar un proyecto de categorías de acuerdo con dichos presupuestos básicos y probar su adecuación a las respuestas dadas en una muestra, recogida de unos 50 ó 100 cuestionarios, a efectos de su corrección conforme al tenor de éstos. También se puede realizar un análisis de contenido de una muestra de los cuestionarios que sirva para establecer las categorías.

3. Las categorías deben ser exhaustivas, es decir, deben existir categorías para todas las clases de respuestas significativas, y categorías generales para clasificar los casos que no quepan en las categorías particulares.

4. Las categorías deben ser excluyentes entre sí, de tal modo que no quepa en ningún caso la posibilidad de inclusión de una respuesta en dos categorías distintas.

5. Su enunciado ha de realizarse del modo más claro posible, de modo que no puedan darse diversas interpretaciones, al ser fácilmente comprensibles de la misma manera por todos.

6. Las categorías deben formarse de manera que hagan referencia a una sola dimensión y aparezcan ordenadas de la manera más homogénea, uniforme y lógica posible. Se debe procurar de manera especial que no sean demasiado generales, sino tan detalladas como sea exigido por los fines de la investigación.

7. En la clasificación de las respuestas en las categorías se ha de dar preferencia a su sentido respecto a su tenor literal.

8. Se ha de procurar que no haya que clasificar una proporción estimable de respuestas en categorías de tipo genérico. En concreto, se estima que la categoría "otros" en ningún caso debe absorber más del 5 por 100 de las respuestas. (Ver ejercicio 150)

FORMACION Y LECTURA DE LAS TABLAS

La codificación general de todos los cuestionarios nos faculta para saber la suma o total de todos los datos de la encuesta referentes a idéntico código.

Por ejemplo, si en el caso de tabulación manual pasáramos una aguja por el orificio 16, supuesto que éste fuera el código de estudios primarios solamente, de todas las fichas nos caerían las que tuvieran cortado este borde por corresponder a los individuos con estudios primarios solamente. Contando todas estas fichas obtendríamos el total de los encuestados que sólo tienen estudios primarios. En la tabulación electrónica esto nos lo dirían las computadoras.

La codificación hace las veces de regla o pauta para la tabulación de los datos. Nos suministra ante todo la suma o total de todos los datos de los cuestionarios o documentos de observación de idéntico código, es decir, del mismo tipo. Existe, pues la necesidad de ordenar estos totales y presentarlos de forma sistemática, lo que se realiza en las tablas.

Estas no son otra cosa que la disposición conjunta y ordenada de las sumas o totales obtenidos en la tabulación de los datos, referentes a las categorías o dimensiones de una variable o de varias relacionadas entre sí.

La formación de las tablas, que mediante los ordenadores electrónicos se puede realizar mecánicamente, constituye un elemento básico de toda investigación empírica. Sistematiza sus resultados cuantitativos y ofrece una visión numérica, sintética y global, del fenómeno estudiado y de las relaciones entre sus distintos aspectos. En ella, en fin, culmina y se concreta definitivamente la fase clasificatoria de la investigación sociológica.

LA EXPLOTACION DEL CUESTIONARIO

La base para realizar la tabulación es lo que se llama la explotación del cuestionario.

Según señala C. Javeau (1971, p. 139), cuando los cuestionarios han sido codificados y pasados a fichas perforadas es preciso *explotarlos*, es decir, es preciso deducir los resultados interesantes que se inscriben en el cuadro definido de las hipótesis de trabajo.

Mediante la explotación de cuestionario se trata principalmente de poner en relación las respuestas a dos o más preguntas diferentes. Esta operación consiste en ver en qué medida dos o más conjuntos de respuestas están ligados entre sí.

Con este fin, se prepara el *plan de explotación*. De hecho, se trata de indicar qué variables deben ser puestas en relación con otras: en otros términos, es preciso indicar qué columnas de la ficha conviene cruzar con otras columnas.

Este plan puede consistir bien en un cuadro general de 80 filas y 80 columnas en la que se designa en relación a cada columna, las otras columnas con las que se debe cruzar, bien en cuadros específicos para cada pregunta o variables relacionadas. Es una operación importante que se debe realizar con sumo esmero.

Las tablas plantean dos problemas prácticos principales: 1) su construcción y 2) su examen, lectura e interpretación.

LA CONSTRUCCION DE LAS TABLAS

La construcción exige la adecuada formulación de sus diversos elementos y resulta de la debida disposición de todos ellos. Estos elementos son: a) el título; b) las columnas; c) la disposición de las variables que la forman; d) las notas en cabeza y marginales que puede presentar; e) las unidades que representan las cifras; f) las columnas de totales; g) los signos empleados y h) las fuentes de donde proceden los datos.

a) El título de la tabla debe contener de manera resumida no sólo la indicación de las variables, sino también la del lugar o comunidad y tiempo a que se refieren los datos de la tabla, así como la mención, en su caso, de circunstancias significativas que afecten o precisen los datos de la tabla.

Conforme se expone en la obra "La documentación y organización de los datos en la investigación sociológica" (Cajas de Ahorro, Madrid, 1969, pág. 111), en el título debe figurar en primer lugar la variable dependiente, después la variable independiente y, por último, la referencia al lugar y año y a otras circunstancias que determinen el sentido de la tabla.

b) En las columnas se puede distinguir la columna matriz de todas las demás. Esta es la que aparece a la izquierda de la tabla y contiene las dimensiones de categorías de la variable dependiente que dominan el contenido de las demás columnas. Estas se refieren a las dimensiones de la variable independiente. Tanto unas como otras se deben ordenar según algún y el mismo criterio.

c) Disposición de las variables. Las tablas, como es sabido, se componen de filas y columnas numéricas. La norma es, aunque no siempre se cumple, colocar las frecuencias referentes a cada categoría o dimensión de las variables independientes verticalmente o en columnas, y los de cada categoría o dimensión de la variable dependiente horizontalmente o en fila. Esto exige, y hay que tenerlo muy en cuenta por la confusión a que da lugar, que las cabeceras de las filas sean las dimensiones en que se divide la variable dependiente y, por el contrario las cabeceras de las columnas sean dimensiones de la variable independiente.

En todo caso, sea ésta o no la disposición de la variable dependiente, se suele indicar el sentido horizontal o vertical de sus valores, si éstos figuran expresados en porcentajes, colocando un 100 por 100 en la primera columna, al pie, si los valores de la variable independiente figuran en sentido vertical y en la primera fila, al final, si figuran en sentido horizontal.

d) Las notas de cabeza y marginales se utilizan para expresar circunstancias sobre el origen, procedencia, finalidad, alcance, validez, etc., de los datos contenidos en la tabla y sirven para una interpretación cabal de ésta.

e) Condición esencial en las tablas es que se exprese claramente, lo que debe hacerse preferentemente en el título, las unidades que representan las cifras que figuran en las tablas: porcentajes, tasas, coeficientes, unidades simples, etc. Si las unidades son porcentajes y sus totales suman más de 100 por 100, por tratarse por ejemplo, de preguntas de respuesta múltiple, debe advertirse así en la tabla, por medio de la nota o indicación correspondiente.

Si las cifras de la tabla son porcentajes, punto fundamental a tener en cuenta en ellos es la llamada regla de Zeisel, según la cual los tantos por ciento se deben hallar y expresar en las tablas en la dirección de la variable independiente o causa, utilizando, por tanto, como base para hallar los tantos por ciento las cifras totales absolutas de la variable independiente. Si los tantos por ciento de una tabla están colocados en dirección indebida y se consignan los números absolutos marginales o totales de que proceden, es posible reconstruir la tabla en la debida dirección, hallando los valores absolutos de cada casilla, colocando éstos en la dirección debida y luego volviendo a hallar los tantos por ciento. (Ver ejercicios 151, 158)

f) Otro de los elementos importantes de las tablas son las columnas de totales. No sólo deben figurar en ellas los totales generales de la tabla, en sentido horizontal y vertical, sino también los totales parciales en el caso de que la tabla se componga de varios grupos. Todas estas columnas de totales, que reciben el nombre de marginales, presentan agrupados o sintetizados los resultados de las tablas para cada una de las variables o características que comprende, y representan los valores de la tabla respecto a una sola variable, si es que la tabla está formada por más de una. (Ver ejercicio 219).

g) También es de interés conocer el significado de los signos más corrientes usados en las tablas, tanto para su correcta utilización como para la debida lectura de aquéllas. Estos son, según la obra citada (Cajas de Ahorro, 1969, pág. 115), los siguientes:

( ): Indica normalmente una base de la que se han extraído porcentajes u otras medidas; en lo posible debe usarse solamente para números absolutos.

*: Se utiliza cuando un porcentaje no es cero, pero menor de 0,5 por 100. Aparece también como llamada a alguna nota o como medida de significación estadística. Debe explicarse qué equivalencia se utiliza en cada caso.

—: Suele indicar el valor nulo o cero de un dato.

...: Suele indicar la no existencia de un dato, sin información, no se ha podido obtener, etc. Se emplea sólo en datos secundarios.

h) La consignación al pie de la tabla de la fuente de donde procede ella o los datos que contiene es una referencia indispensable que no debe faltar en ninguna tabla. Esta mención debe ajustarse a las normas generales admitidas y seguidas en las citas bibliográficas: Autor. Título del estudio. Lugar de edición. Editorial y año. Estas dos últimas referencias se deben sustituir en el caso de las revistas por el nombre de la revista o publicación periódica de que se trate y por la indicación del volumen, número y año del ejemplar en que ha aparecido el trabajo de donde procede la tabla.

Si se tienen en cuenta todas estas indicaciones sobre los elementos de las tablas se podrán construir éstas correctamente sin que falte en ella ninguna indicación importante.

Por último deben tener en cuenta que en las operaciones de formación de las tablas, como en todas las demás, se pueden cometer errores de atribución, olvido de cuestionarios o fichas, etc.

Por ello las tablas deben ser objeto de comprobación aritmética minuciosa respecto a las sumas totales y porcentajes, sobre su concordancia y compatibilidad con resultados similares y de la misma naturaleza y sobre su verosimilidad. (Ver ejercicios 151, 152, 153, 154)

LECTURA DE LAS TABLAS

Respecto a la lectura de las tablas, se debe tener en cuenta, en primer lugar, la regla indicada de Zeisel, según la cual se deben efectuar las comparaciones en la dirección opuesta a la que se han calculado los porcentajes, es decir, si los tantos por ciento se han hallado en dirección vertical interesa, antes que comparar unos con otros en esta dirección, comparar horizontalmente los tantos por ciento que figuran en las distintas columnas en nivel similar, o en la misma líneas. (Ver ejercicios 156 a 158).

Como conjunto de reglas generales a seguir en este punto se transcriben las siguientes:

1. Leer el título cuidadosamente.
2. Leer atentamente las notas o cualquiera otras explicaciones a la cabeza de la tabla
3. Fijarse en la fuente o el origen de donde se ha tomado la tabla.
4. Fijarse en las notas a pie de página si existen.
5. Hallar qué unidades son usadas en la tablas.
6. Fijarse en los promedios totales.
7. Ver la variabilidad que existe entre los datos de la tabla.
8. Mirar cómo están relacionados o asociados los porcentajes en cada uno de los criterios principales de clasificación.
9. Examinar la consistencia de los resultados y la interacción entre ellos.
10. Finalmente, prestar atención a las aberraciones, anormalidades o irregularidades que se observan en los resultados de la tabla. (Ver ejercicios 153, 155, 165 y 166).

CONSTRUCCION DE TIPOLOGIAS

Las operaciones de clasificación, aludidas hasta aquí, que culminan en la confección de las tablas, aunque tienen un carácter esencialmente cuantitativo, se basan y fundan, salvo en las tablas descriptivas simples, en la combinación de dos o más variables. Este cruzamiento de variables tiene la particularidad de que se puede realizar haciendo abstracción de valores numéricos, y tomando en consideración las dimensiones de tipo cualitativo o cuantitativo abstracto que se pueden distinguir en las variables cualitativas.

Esta operación de la combinación de las dimensiones cualitativas de dos o más variables da lugar a aparición de diversos tipos en que se combinan de distinta forma dichas dimensiones y que en conjunto reciben el nombre de tipologías. La afinidad indicada con las tablas numéricas en cuanto resultado también de cruzamiento de variables, justifica que se trate aquí de las tipologías, a pesar de su diferente carácter cualitativo y de su aplicación como instrumento de análisis y de síntesis teóricos.

Respecto a las tipologías se abordan en los ejercicios tres puntos. Su formación y las operaciones llamadas de reducción y sustrucción de tipos.

El modo de operar para la formación de tipologías se puede sintetizar como sigue:

1. Es necesario que las variables elegidas para la formación de la tipología estén relacionadas entre sí.

2. Elegidas las variables, se coloca una en sentido horizontal y la otra en sentido vertical.

3. Se subdivide el espacio horizontal o vertical dedicado a cada variable, según las dimensiones o grados que se distingan o hayan distinguido en aquéllas.

4. Por medio de líneas horizontales y verticales paralelas a las variables se construye un cuadrado o rectángulo, que se subdivide a su vez por líneas paralelas internas también verticales y horizontales en rectángulos o cuadrados pequeños, casillas, según los grados que se hayan distinguido en cada variable. Las casillas resultantes se denominan con los nombres de las dos variables y sus dimensiones correspondientes relacionadas en dicha casilla, aunque es corriente asignarlas un apelativo que sintetice o englobe las dos características cruzadas.

Si las variables que intervienen en la tipología son tres, de dos dimensiones cada una, entonces dos de las variables, las independientes, se colocan en sentido horizontal. Si fueran cuatro las variables, también de dos dimensiones se colocarían, de modo similar al acabado de exponer, dos horizontalmente y otras dos verticalmente. (Ver ejercicios 159, 160)

La reducción de tipologías consiste simplemente en transformar una tipología de X número de casillas en otra de menor número mediante la refundición de las dimensiones de sus variables, en el caso de que dichas dimensiones sean más de dos. (Ver ejercicios 161, 162)

La sustrucción, por su parte, consiste en hallar, dada una serie de tipos, las posibles variables y dimensiones a que pueden responder, construyendo a base de ellos la tipología correspondiente. Se basa en el análisis conjunto de los tipos para descubrir en ellos sus elementos comunes o variables y las dimensiones de estos elementos. (Ver ejercicios 163, 164)

EJERCICIOS

149

En una encuesta verificada a veraneantes en Torremolinos en el cuestionario aplicado figuran las siguientes en las que se indican con una cruz las contestaciones dadas por un encuestado.

20. ¿Tiene usted hijos?

Sí. X
No.
No contesta.

21. ¿Cuántos hijos tiene?

1.
2.
3. X
4.
5.
6 o más.
No contesta.
No preguntado.

22. ¿Cuántos años en total ha estudiado usted, incluida la enseñanza primaria?

.... 14. X.
No contesta.

23 ¿Es usted español?

Sí.
No. X
No contesta.

24. Si es usted español ¿cuál es la región de su procedencia?

Castilla la Vieja.
Castilla la Nueva.
León.
Aragón.
Cataluña.
Valencia.
Baleares.
Provincias Vascongadas.
Navarra.
Asturias.
Galicia.
Extremadura.
Andalucía.
Canarias.
No contesta.
No preguntado. X

Se pide codificar estas preguntas, teniendo en cuenta que aparecen seguidas en el cuestionario y a la pregunta anterior a efectos de tabulación electrónica se le ha asignado la columna 25, y que se sigue, en éste como en los demás ejercicios, el procedimiento de consignar la clave de codificación de cada pregunta a su margen derecho.

RESPUESTA

20. ¿Tiene usted hijos?

1. Sí.
2. No.
9. No contesta.

1
Col. 26

21. ¿Cuántos hijos tiene?

0. No preguntado.
1. Uno.
2. Dos.
3. Tres.
4. Cuatro.
5. Cinco.
6. Seis o más.
9. No contesta.

3
Col. 27

22. ¿Cuántos años ha estudiado usted, incluida la enseñanza primaria?

Catorce.

1,4
Cols. 28 y 29

23. ¿Es usted español?

1. Sí
2. No.
3. No contesta.

2
Col. 29

24. Si es usted español ¿en qué región vive habitualmente?

00. No pregunta.
01. Castilla la Vieja.
02. Castilla la Nueva.
03. León
04. Aragón.
05. Cataluña.
06. Valencia.
07. Baleares.
08. Vascongadas.
09. Navarra.
10. Asturias.
11. Galicia.
12. Extremadura.
13. Canarias.
19. No contesta.

00
Cols. 30 y 31

Como se ve, se ha dedicado siempre el número 0 a la no pregunta y el 9 a la no contestación, pues como son categorías que se repiten en todas de la misma forma en las preguntas se acostumbra a dedicarlas siempre el mismo dígito.

En la pregunta 22, como la respuesta es numérica, se destinan una columna para cada número que componga la cifra de que se trate. Asimismo, en la pregunta 24, al tener más de doce categorías se destinan también dos columnas.

150

En una encuesta realizada en la República Federal Alemana, se dieron a esta pregunta abierta, "Si usted hace el favor de pensar ahora en el futuro próximo, quiero decir, en el año que viene, ¿cuál es su mayor esperanza para ese año?", las respuestas, entre otras, que se indican a continuación: 1. "Visitar a mis hijos". 2. "Que siga habiendo paz y tranquilidad en el país". 3. "Que yo obtenga pronto una vivienda mayor". 4. "Ganar tanto en las quinielas que no necesite trabajar fuera de casa". 5. "Mejores condiciones meteorológicas para una buena cosecha". 6. "Un viaje a Viena". 7. "Salud". 8. "Tener de una vez vacaciones con mi marido". 9. "Que no tengamos guerra". 10. "Que mejore mi salud". 11. "Que pueda celebrar todavía mis bodas de plata". 12. "Pagar mis deudas por la nueva casa". 13. "Un cambio de trabajo para el extranjero". 14. "Quisiera casarme". 15. "No tengo ningún deseo". 16. "Que mi equipo suba a primera". 17. "Un primogénito sano". 18. "Que sea incluido en la relación de funcionarios". 19. "Ser admitido en el club social tal". 20. "Llegar a relacionarme con la buena sociedad".

Inténtese categorizar estas respuestas a efectos de su codificación y realizar ésta teniendo en cuenta que se ha asignado a la pregunta anterior la columna 18:

1. En el caso de que el fin pretendido sea conocer la importancia de las expectativas de tipo económico respecto a las demás.

2. Supuesto que se quiere tener un conocimiento general de sus formas principales.

RESPUESTA

En el primer caso, como se pretende conocer sólo la importancia de las expectativas económicas respecto a las demás, bastará formar con ellas las dos categorías siguientes:

1. Esperanzas de tipo económico.
2. Otras esperanzas.
3. Ningún deseo.
9. Sin respuesta.

En el segundo caso las categorías podrían ser las siguientes:

0. Ningún deseo.
1. Esperanzas y deseos de salud.
2. Esperanzas de tipo familiar.
3. Esperanzas de tipo económico.
4. Esperanzas de tipo profesional.
5. Esperanzas sobre la vivienda y el hogar.
6. Esperanzas de viajes y vacaciones.

7. Esperanzas de tipo político y social.
8. Otros deseos y esperanzas.
9. Sin respuesta.

De acuerdo con estas categorías, las preguntas dadas se codificarían en la columna 19, y en el caso de la primera forma de categorización con los dígitos 1, las respuestas del enunciado 4, 5, 12; con el 2, las 1, 2, 3, 6, 7, 8, 9, 10, 11, 13, 14, 16, 17, 18, 19, 20; y con el 3 la pregunta 15. En cuanto a la segunda forma de categorización se codificarían todas las respuestas en la misma columna 19 y en los dígitos 0, la respuesta del enunciado 15; en el dígito 1, las 7, 10, 11; en el 2, las 1, 14, 17; en el 3, las 4, 5 y 12; en el 4, la 13 y la 18; en el 5, la 3; en el 6, la 6 y la 8; en el 7, la 2 y la 9; en el 8, la 20 y la 19; en el 9, ninguna, consignándose debajo de los dígitos correspondientes, como siempre, el número de la columna a perforar.

(Basado en E. Noelle: *Encuestas en la sociedad de masas.* Alianza Editorial. Madrid, 1970, pág. 232.)

151

Dadas las siguientes tablas de resultados de investigaciones sociológicas realizadas, se pide: a) señalar la variable dependiente y la independiente; b) indicar en qué están expresadas las unidades de las tablas; c) titularlas; d) expresar la significación de los marginales o totales, respecto de los cuales se advierte que las cifras consignadas en las tablas son supuestas, y e) formular el sentido en que deben hacerse las comparaciones.

El lugar es Madrid y el año 1970.

A) TABLA PRIMERA

Frecuencia con que van al cine	*Devotos*	*Menos devotos*	*Total devotos y no devotos*
Por lo menos una vez por semana	31	49	38
De vez en cuando	42	36	40
Nunca ve películas	27	15	22
	100 %		
N (Número total de casos)	(520)	(334)	(854)

RESPUESTA

a) La variable dependiente es la "frecuencia con que van al cine", y la variable independiente, en la dirección de la cual figuran hallados los porcentajes, la devoción religiosa.

b) Las unidades de la tabla expresan porcentajes, como lo indica que sumen cien verticalmente todas las columnas y que las cifras absolutas o los números totales de casos figuren entre paréntesis al margen de la tabla, lo que quiere decir que constituyen la base de los porcentajes de la tabla hallados.

c) El título procedente sería: "Frecuencia de ir al cine en porcentajes según la devoción religiosa. Madrid, 1970".

d) Los marginales o totales expresan las frecuencias de la variable dependiente haciendo abstracción de la variable independiente. En concreto, indican que la población estudiada, independientemente de toda otra consideración, un 38 por 100 va por lo menos una vez al cine por semana; un 40 por 100, de vez en cuando, y un 22 por 100 nunca ven películas.

e) Según la regla de Zeisel, las comparaciones y lectura de las tablas se deben hacer en sentido contrario al que figuran hallados los porcentajes. Como en este caso figuran hallados en sentido vertical, es preciso, pues, realizarla en sentido horizontal, o sea, en lugar de decir de los devotos un 31 por 100 van por lo menos una vez por semana al cine, un 42 por 100 le ven de cuando en cuando y un 27 por 100 nunca, lo que tiene menos significación, es preciso leer que van por lo menos una vez al cine por semana de los devotos un 31 por 100 y de los menos devotos un 49 por 100, y así sucesivamente, de donde se deriva que, a menos devoción, la frecuencia de ir al cine es mayor.

B) TABLA SEGUNDA

	Interés político		
Educación	*Alto*	*Bajo*	*Núm de casos*
Primaria	15	85	100 % (525)
Secundaria	28	72	(280)
Universitaria	40	60	(175)
Total marginal	22	78	(930)

RESPUESTA

a) La variable dependiente es el interés político, y la variable independiente, la educación.

b) Las unidades de la tabla expresan porcentajes por los mismos motivos que en la tabla anterior.

c) El título correspondiente es: "Interés político en porcentajes según nivel de educación en Madrid (1970)."

d) Los marginales totales tienen también el mismo significado que en la tabla anterior. Indican que en la población estudiada en un 22 por 100 el interés político es alto, y en un 78 por 100, bajo.

e) Como en esta tabla los porcentajes figuran hallados en dirección horizontal, es preciso realizar las comparaciones y lectura en sentido vertical. Por tanto, es preciso leer que tienen un interés político alto un 15 por 100 de los que tienen estudios primarios, un 28 por 100 de los que los tienen secundarios y un 40 por 100

de los que los tienen universitarios, y así en relación al interés político bajo, de donde se deriva que a mayor nivel de educación el interés político es mayor.

152

En una encuesta realizada sobre los motivos de compra de una crema para la cara, de la tabulación realizada se ha obtenido la siguiente tabla, que figura en Goode y Hatt (1952, pág. 522).

Motivos de compra de una crema para la cara:

1. Recomendaciones	*28*
2. Beneficio de la piel	*21*
3. Publicidad radiofónica	*18*
4. Visto en la tienda	*15*
5. Buen precio	*10*
6. Buen perfume	*8*
7. Condiciones especiales de la piel	*7*
Suma total de porcentajes	*107*

Se pide:

a) Señalar las variables dependiente e independiente de la tabla.

b) Titularla.

c) Indicar las unidades que la forman.

d) Indicar, en fin, si los motivos están bien ordenados y cómo se podría mejorar su ordenación, presentando la tabla, en su caso, según ella.

RESPUESTA

a) Se trata de una tabla de una sola variable, los motivos de la compra.

b) Las unidades que la forman, como se indica en el total, son porcentajes.

c) El título sería, por tanto: "Motivos de compra de la crema X para la cara en porcentajes", más lugar y fecha, que no se indican.

d) El examen de la tabla muestra un desorden en la presentación de los motivos, lo que dificulta su comprensión. Para mejorar ésta es necesario articular la tabla agrupando los motivos en dos o tres categorías según su condición.

Los motivos se pueden clasificar en personales, 2, 7; referentes al producto, 5, 6, y referentes a las influencias externas, 1, 3 y 4.

De acuerdo con esta clasificación, la tabla sería la siguiente:

MOTIVOS DE COMPRA DE UNA CREMA PARA LA CARA
EN PORCENTAJES

	Porcentajes	
Personales:		
Beneficio de la piel	21	
Condiciones especiales de la piel	7	28
Referentes al producto:		
Buen precio	10	
Buen perfume	8	18
Referentes al modo de su conocimiento:		
Recomendaciones	28	
Publicidad radio	18	
Visto tienda	15	61
SUMA TOTAL DE PORCENTAJES		107

Como se ve la claridad de la tabla es mucho mayor presentándola de esta forma, lo que prueba la necesidad, en casos similares, de presentar ordenados lógicamente los conceptos.

153

A) *Dada la siguiente tabla:*

	Porcentajes de delincuencia. Grado de rigidez de la disciplina materna		
Sexo	*Rígida* %	*Justamente tolerante* %	*Muy tolerante* %
Muchachos	30 (100)	30 (160)	40 (100)
Muchachas	20 (120)	30 (170)	40 (100)

Se pide determinar:

a) ¿Cuál es el número total de casos de la tabla?

b) ¿Cuántos muchachos son delincuentes?

c) ¿Cuántas muchachas no delincuentes proceden de familias donde la disciplina de la madre es muy tolerante?

d) ¿Cuántas variables aparecen en la tabla y cómo se titularía?

RESPUESTA

a) Para hallar el número total de casos de la tabla basta sumar los números absolutos que figuran entre paréntesis en ella: 100, 120, 160, 170, 100, y 100, que hacen un total de 750.

b) El número de jóvenes delincuentes se obtiene hallando los tantos por ciento de las cifras absolutas de jóvenes, tantos por ciento y cifras absolutas que figuran en la fila de la tabla correspondiente a los muchachos. Es, pues, la suma del 30 por 100 de 100, el 30 por 100 de 160 y el 40 por 100 de 100, o sea, 30, 48 y 40, que sumados, dan un total de jóvenes delincuentes de 118.

c) Las cifras que corresponden en la tabla a las chicas delincuentes procedentes de casas donde la disciplina de la madre es muy tolerante es el 40 por 100 y una cifra absoluta de 100. Luego el número de chicas no delincuentes de esta categoría será la diferencia entre el total de chicas de ella, 100, y el 40 por 100 de 100, 40, que son delincuentes, o sea, 60.

d) Las variables de la tabla son tres: Una dependiente, la delincuencia, y dos independientes o explicativas, el sexo y el grado de rigidez de la disciplina materna. El título de la tabla sería pues: "Delincuencia juvenil en porcentajes según el sexo y el grado de rigidez de la disciplina materna."

B) *Dada la tabla del ejercicio anterior, la que figura expresada, como se habrá observado, en forma reducida, es decir, consignando únicamente los porcentajes de delincuentes y omitiendo los de no delincuentes, se pide formar la tabla completa.*

RESPUESTA

Delincuencia juvenil según el sexo y el grado de rigidez de la disciplina materna (en porcentajes).

	Grado de rigidez de la disciplina materna					
	Rígida		*Justamente tolerante*		*Muy tolerante*	
SEXO	*Delinc.*	*No delinc.*	*Delinc.*	*No delinc.*	*Delinc.*	*No delinc.*
Chicos	30	70 (100)	30	70 (160)	40	60 (100)
Chicas	20	80 (120)	30	70 (170)	40	60 (100)

Como se ve, la operación consiste simplemente en hallar los porcentajes omitidos de jóvenes no delincuentes, que se obtienen deduciendo del porcentaje total, 100 por 100, el porcentaje dado de jóvenes delincuentes.

154

Dada la siguiente tabla:

NIVEL EDUCATIVO, SEGUN EDAD Y CLASE SOCIAL
(en porcentajes)

	Jóvenes (clase social)			*Viejos (clase social)*		
Nivel educativo	*Alta*	*Media*	*Baja*	*Alta*	*Media*	*Baja*
Bachillerato superior o más	85	65	30	70	50	10
Menos de bachillerato superior	15	35	70	30	50	90
	100 % (100)	(200)	(300)	(50)	(120)	(260)

Se pide reducir la tabla omitiendo la dimensión, menos de bachillerato superior, titularla e indicar las variables que la forman reducida y sin reducir:

RESPUESTA

Título: Porcentajes de nivel educativo Bachiller superior o más, según edad y clase social.

	Edad	
CLASE SOCIAL	*Jóvenes*	*Viejos*
Alta	85 (100)	70 (50)
Media	65 (200)	50 (120)
Baja	30 (300)	10 (260)

En ambas tablas las variables son tres, y las mismas: una dependiente, nivel de estudios y dos independientes, edad y clase social.

155

Dada la siguiente tabla, basada en Zeisel, 1962, pág. 97:

Porcentajes de individuos que prefieren las radioemisiones extranjeras de su país de origen, según nación, generación y transmisión de las tradiciones patrias:

	Alemanes			*Polacos*			*Italianos*		
	Generación			*Generación*			*Generación*		
	1ª.	*2ª.*	*Total*	*1ª.*	*2ª.*	*Total*	*1ª.*	*2ª.*	*Total*
Enseñan a los hijos las tradiciones de la madre patria	20	0	12	57	14	49	60	6	41
No se las enseñan	8	0	4	32	0	14	41	4	18
Total (%)	12	0	5	47	3	29	53	5	30

Se pide formar con los totales o marginales, e interpretarlas, las tablas de preferencia de las radioemisiones extranjeras según la nacionalidad, y de influencia de la generación en la preferencia de dichas emisiones. Formar e interpretar asimismo con los porcentajes no totales de las tablas las tablas de la influencia de la tradición en la preferencia de las emisiones en cuestión, respecto a la primera y segunda generación separadamente. Señalar en cada caso el número de variables de las tablas formadas.

RESPUESTA

Primera tabla: Porcentajes de preferencia de las radioemisiones extranjeras del país de origen, según la nacionalidad:

	%
Alemanes	5
Polacos	29
Italianos	30

El cuadro indica que son los polacos y los italianos quienes tienen una mayor preferencia por las emisiones de su país. Las variables son dos en este caso: la preferencia de las emisiones y la nacionalidad.

Segunda tabla: Porcentajes de preferencia de las radioemisiones del país de origen según la nacionalidad y la generación:

	Alemanes	*Polacos*	*Italianos*
Primera generación	12	47	53
Segunda generación	0	3	5

Las variables de esta tabla son tres: la preferencia de las radioemisiones, la generación y la nacionalidad. El cuadro dice que, dentro de una preferencia mayor en los polacos e italianos que los alemanes, en la segunda generación los porcentajes de dicha preferencia son casi insignificantes en las tres nacionalidades. Por tanto, dentro de cada nacionalidad, la proporción de los que prefieren las radioemisiones del país de origen depende fuertemente de la generación de que se trate.

Tercera tabla: Porcentajes de preferencia de las radioemisiones del país de origen según la nacionalidad y la tradición, y con referencia a la primera generación.

	Alemanes	*Polacos*	*Italianos*
Recibieron la tradición	20	57	60
No la recibieron	8	32	41

En la tabla se relacionan cuatro variables: la preferencia de las radioemisiones, la nacionalidad, la recepción de las tradiciones y, además, la generación, en cuanto todos los porcentajes se refieren a la primera generación. Indica que la preferencia por las emisiones del país de origen dentro de la primera generación es, en los alemanes que recibieron la tradición, algo superior al doble que en los que no la recibieron; en los polacos, algo inferior al doble, y en los italianos, superior en un 50 por 100.

Cuarta tabla: Porcentajes de preferencia de las radioemisiones del país de origen según la nacionalidad y la tradición y con referencia a la segunda generación:

	Alemanes	*Polacos*	*Italianos*
Recibieron la tradición	0	14	6
No la recibieron	0	0	4

Las variables son las mismas que en la tercera tabla: su interpretación señala que en los alemanes la influencia de la tradición es nula con referencia a la segunda generación; que en los polacos es muy intensa, pues mientras en los que no recibieron la tradición la preferencia es 0, en los que la recibieron esta preferencia se eleva a un 14 por 100; en los italianos, la preferencia de los que recibieron la tradición es como en el caso anterior, un 50 por 100 superior a la de los que no la recibieron.

Este ejercicio muestra la utilidad de los marginales para leer e interpretar las tablas sobre todo si son complicados o de muchas variables como ésta. Zeisel, en la obra citada, recomienda utilizar en estos casos este procedimiento u otro similar, ante la dificultad de hacer la interpretación de la tabla como en las sencillas, mediante comparaciones de las cifras que figuran en ella horizontal o verticalmente.

156

Dada la siguiente tabla:

PORCENTAJE DE MAESTROS EN HUELGA

	Actitud hacia los Sindicatos	
Clase de origen	*Favorable* %	*Desfavorable* %
Baja	70 (70)	30 (40)
Media	70 (70)	30 (120)

Se pide determinar:

A) ¿Qué proporción de maestros procedentes de familia de la clase baja están en huelga?

B) ¿Qué proporción de maestros procedentes de la clase media están en huelga?

C) Indicar si existe correlación entre la clase de origen y la actitud hacia el Sindicato.

D) Explicar razonadamente si la clase de origen es una causa de huelga espúrea o falsa.

E) Explicar razonadamente si la actitud hacia los Sindicatos es una causa de huelga espúrea.

F) ¿Qué nos dice la tabla?

RESPUESTA

A. Para hallar la proporción de maestros en huelga que proceden de familias de clase baja hay que sumar el número absoluto en huelga de los de dicha procedencia que tienen una actitud favorable y desfavorable al sindicato, o sea, sumar el 70 por 100 de 70 y el 30 por 100 de 40, 49 y 12, respectivamente, que hacen un total de 61, y obtener el porcentaje que representa esta suma respecto al total de maestros de la clase baja de la tabla: 70 más 40, igual 110.

$$\text{Proporción de maestros de la clase baja en huelga} = \frac{61 \times 100}{110} = 55$$

B. Realizando las mismas operaciones respecto a los maestros procedentes de familia de la clase media se obtiene 70 por 100 de 70, 49; más 30 por 100 de 120, 36 igual a 85. Total de maestros de la clase media: 120 + 70 = 190.

$$\text{Proporción de maestros de la clase media en huelga} = \frac{85 \times 100}{190} = 44$$

C. De los datos de la tabla no se deriva que exista correlación entre la clase de origen y la actitud hacia los sindicatos porque los porcentajes de maestros en huelga de la clase baja que tienen una actitud favorable y desfavorable a los sindicatos 70 y 30 por 100, respectivamente, son idénticos a los porcentajes correspondientes de maestros procedentes de la clase media.

D. De la tabla no se deriva que la clase social sea una causa espúrea o falsa de huelga, ya que según ella existe una proporción mayor de maestros en huelga procedentes de la clase baja, el 55 por 100, que de la clase media, 44 por 100.

E. Más claro aparece, según la tabla, que la actitud hacia los sindicatos no es una causa espúrea o falsa de huelga, en cuanto es mucho mayor, 70 por 100, el porcentaje total de los maestros que ha ido a la huelga, favorables a los sindicatos que el de los desfavorables a éstos, 30 por 100, que también han ido a la huelga.

F. La tabla, en síntesis, nos muestra que existe una correlación fuerte entre la actitud favorable o desfavorable a los sindicatos y la declaración en huelga y una correlación menos fuerte entre la clase baja y media y dicha declaración en huelga, sin que de la tabla se derive correlación alguna entre la clase social y la actitud hacia el sindicato.

157

Indicar razonadamente si las tablas que siguen confirman o no las proposiciones que se exponen:

A) Las chicas obtendrán probablemente mejores notas en la Facultad X.

Sexo	*Estudiantes con nota media de notable o superior (%)*
Muchachas	30 (1.000)
Muchachos	30 (1.000)

RESPUESTA

La tabla no confirma la proposición. Dado que los porcentajes de obtención de notable o más son iguales en los chicos que en las chicas, la tabla no autoriza a afirmar que las chicas obtendrán probablemente mejores calificaciones medias que los chicos.

B) Los chicos destacarán probablemente en ciencias matemáticas.

Sexo	*Destacan en ciencias o matemáticas* %
Chicos	30 (1.000)
Chicas	10 (1.000)

RESPUESTA

Sí confirma la proposición. Indica que destacan en ciencias o matemáticas un 30 por 100 de los chicos frente a sólo un 10 por 100 de las chicas. Por tanto, la tabla da pie para afirmar que los chicos destacarán probablemente en relación a las chicas.

C) Los maestros de las escuelas elementales apoyarán probablemente menos la huelga que los maestros de las escuelas secundarias, porque la mayor parte de los maestros de las escuelas elementales son mujeres y éstas apoyan probablemente menos la huelga que los hombres.

Tipo de escuela	*Maestros que apoyan la huelga*	
	Hombres %	*Mujeres* %
Elemental	30 (100)	30 (900)
Secundaria	60 (400)	60 (400)

RESPUESTA

No lo confirma. De la tabla se deriva que el porcentaje de los maestros que apoyan la huelga es igual al de maestras, tanto en las escuelas elementales como en las secundarias. Hay, sí, un mayor apoyo a la huelga en las escuelas secundarias que en las elementales, pero de la tabla no se deriva que influya en ello algo el sexo o el mayor número de mujeres en las escuelas elementales.

158

Examinar cada una de las siguientes tablas y determinar si los porcentajes en ellas están establecidos correctamente. Si no es así, volver a hallar los porcentajes en la dirección debida. Expresar también el título de las tablas y efectuar su lectura.

A) Tipos de familia	*Delincuencia juvenil* %
Convivencia con los dos padres	6 (1.000)
Convivencia con un solo padre	20 (200)
No convivencia con padres	25 (50)

RESPUESTA

Se trata de ejercicios de aplicación de la regla de Zeisel según la cual, como es sabido, los porcentajes en la tabla se deben hallar en la dirección de la variable independiente o causa. Por tanto, para ver si están bien dispuestos los porcentajes es preciso determinar antes cuál es en principio o normalmente la variable independiente.

En este caso, la variable independiente o causa es el tipo de familia. Es de ella de la que depende en principio el mayor o menor porcentaje de delincuencia juvenil y la que es anterior en el tiempo y puede influir en la delincuencia. Los porcentajes están hallados en la dirección de esta variable independiente, luego está bien dispuesta la tabla en cuestión y no necesita modificación. Según ella, el porcentaje de delincuencia es sólo del 6 por 100 en los casos de convivencia con los dos padres, mientras que se eleva al 20 y 25 por 100, respectivamente, si se convive con un solo padre o sin padres.

El título de la tabla sería: "Porcentaje de delincuencia juvenil según el tipo de familia."

B) Voto a favor de	*Clase trabajadora* %
Repúblicanos	40 (800)
Demócratas	60 (1.000)

RESPUESTA

En esta tabla la variable independiente es la clase social, en cuanto es anterior en el tiempo al voto y éste puede depender de ella. Pero los porcentajes figuran hallados en la dirección de la variable dependiente, voto, luego están mal determinados y es preciso cambiar su disposición, situando los valores de la variable independiente, clase social, horizontalmente, y los de la variable dependiente, voto, verticalmente.

Para ello, como se sabe por la tabla inicial, el total de los que votaron a los demócratas y a los republicanos, 1.000 y 800, respectivamente, y que de ellos pertenecían a la clase trabajadora un 60 por 100 de los demócratas y un 40 por 100 de los republicanos, luego hallando estos tantos por ciento de dichos totales obtendremos en cifras absolutas el número de trabajadores que votaron a los demócratas y a los republicanos y por sustracción con los mismos totales el de los no trabajadores que votaron a los partidos indicados. Con estas cifras absolutas formaremos la tabla poniendo en la dirección debida indicada ambas variables, después de lo cual sólo resta hallar los porcentajes en la tabla formada.

La tabla de cifras absolutas sería la siguiente:

	Clase social		
	Trabajadores	*No trabajadores*	
Voto a favor de:			
Demócratas	600	400	(1.000)
Republicanos	320	480	(800)
TOTALES	920	880	(1.800)

La tabla de porcentajes sería, obtenidos éstos en la dirección vertical de la variable independiente, y no como antes, que estaban en dirección horizontal; si bien hay que tener en cuenta que en la tabla faltaban la columna correspondiente a los no trabajadores.

	Clase social	
	Trabajadores (%)	*No trabajadores* (%)
Voto a favor de:		
Demócratas	65	45
Republicanos	35	55
	100	100
(N.) Número de casos ...	(920)	(880)

Realizando la lectura, según la regla de Zeisel, en sentido horizontal, la tabla nos dice que votan a favor de los demócratas un 65 por 100 de los trabajadores y un 45 por 100 de los no trabajadores; y a favor de los republicanos un 35 por 100 de los trabajadores y un 55 por 100 de los no trabajadores. Según la tabla, existe, pues, una relación entre la clase social y el voto: los trabajadores votan en mayor proporción a los demócratas, y los no trabajadores a los republicanos.

C) Drogas

Frecuencia en el uso de drogas	*Muchachas* %
Han fumado marihuana por lo menos una vez	40 (600)
Nunca han fumado marihuana	40 (200)

RESPUESTA

En esta tabla hay dos variables; el uso de drogas y el sexo (chicos y chicas, pero falta la columna de chicos en la tabla). La variable independiente es el sexo, es la anterior en el tiempo al uso de drogas y de la que puede depender ésta. Pero los porcentajes están hallados en la dirección, como en el caso anterior de la variable dependiente, en sentido horizontal; luego están mal establecidos y es preciso recomponer la tabla en la debida dirección procediendo como en el caso anterior. La tabla de las cifras absolutas en la dirección debida sería:

	Sexo		
	Chicas	*Chicos*	
Frecuencia en el uso de drogas			
Han fumado marihuana por lo menos una vez	240	360	600
Nunca han fumado marihuana	80	120	200
TOTALES	320	480	800

La tabla de porcentajes, obtenidos como antes en la dirección vertical de la variable independiente, y no como antes, que estaban en dirección horizontal, sería la siguiente:

	Sexo	
	Chicas %	*Chicos* %
Frecuencia en el uso de drogas		
Han fumado marihuana por lo menos una vez ...	75	75
Nunca han fumado marihuana	25	25
	100	
N = Número de casos	(320)	(480)

La lectura de la tabla en sentido horizontal nos dice que han fumado marihuana por lo menos una vez el 75 por 100 de las chicas e igualmente el 75 por 100 de los chicos, y que nunca han fumado marihuana un 25 por 100 de ambos. No existe, pues, según la tabla, relación entre el sexo y la frecuencia en el uso de las drogas.

(Los enunciados de los ejercicios 217 a 220, ambos inclusive, proceden de la obra de S. Cole, *Sociological Method.*)

159

Dadas las dos variables producción y consumo, distinguir en ellas dos dimensiones cuantitativas abstractas y formar una tipología de los niveles de desarrollo económico de las sociedades. Exponer la denominación de los tipos de cada casilla y procurar darles también un nombre figurado, comprensivo o técnico.

RESPUESTA

Las dos dimensiones que se pueden distinguir en ambas variables son: Producción: alta, baja; consumo: alto, bajo.

Tipología:

Consumo	*Producción*	
	Alta	*Baja*
Alto	Producción alta. Consumo alto. Sociedades de desarrollo alto.	Producción baja. Consumo alto. Sociedades de economía regresiva
Bajo	Producción alta. Consumo bajo. Sociedades en desarrollo.	Producción baja. Consumo bajo. Sociedades estancadas.

160

Dadas las dos variables: orientación productiva y deseos de consumo, distinguir, como en el caso anterior, en ellas dos dimensiones cuantitativas y formar una tipología de las diferentes actitudes sociales ante el desarrollo.

Exponer la denominación de los tipos de cada casilla y procurar darles también un nombre figurado, comprensivo o técnico.

RESPUESTA

Las dos dimensiones que se pueden distinguir en ambas variables son: Orientación productiva: alta, baja; deseo de consumo: alto, bajo.

Este ejercicio y el anterior se basan en Alan Wells: *Toward an empirically grounded theory of developpment.* "The British Journal of Sociology", vol. 23, 3, 1972.)

Tipología:

Deseo de consumo	*Orientación productiva*	
	Alta	*Baja*
Alto	Orientación productiva alta, deseo de consumo alto. Actitud hedonista.	Orientación productiva baja, deseo de consumo alto. Actitud parásita.
Bajo	Orientación productiva alta, deseo de consumo bajo. Actitud ascética.	Orientación productiva baja, deseo de consumo bajo. Actitud tradicional.

161

Dada la tipología que se expone a continuación, se pide efectuar la reducción de sus categorías o tipos y describirlas, señalando previamente las variables y sus dimensiones.

Grado de interés político	*Simpatías políticas habituales*		
	Republicanos	*Demócratas*	*Independientes*
Alto			
Medio			
Bajo			

RESPUESTA

Variables: 1. Grado de interés político con tres dimensiones: alto, medio y débil. 2. Simpatías políticas habituales, con otras tres dimensiones: republicanos, demócratas, independientes.

En este caso, para efectuar la reducción pedida se pueden unir, por una parte, las dos categorías o dimensiones de republicanos y demócratas, pues son similares, y, por otra, se puede suprimir la categoría media en el grado de interés político y dejar las dos extremas, alto y débil. De esta forma la tipología queda reducida a cuatro tipos. De ellos, el primero se podría llamar de los partisanos, o de las personas con simpatías políticas habituales definidas y alto grado de interés político. El segundo, el de los independientes, no definidos en relación a sus simpatías políticas y con alto grado de interés político. El tercero sería el de los habituados o personas de simpatías políticas definidas y con un interés político débil. Por último, el cuarto tipo sería el de los apáticos o personas con simpatías políticas indefinidas o independientes y bajo grado de interés político.

Grado de interés político	*Simpatías políticas*	
	Republicanos y demócratas	*Independientes*
Alto	Partisanos.	Independientes.
Débil	Habituados.	Apáticos.

(Basado en A. Barton: *Le concept d'espace d'attibuts en Sociologie.* "Le Vocabulaire des Sciences Sociales". Mounton, 1965.)

162

Dada la siguiente tipología:

Conducta moral	*Práctica religiosa*		
	Alta	*Media*	*Baja*
Buena			
Mediocre			
Mala			

Se pide efectuar su reducción designando los tipos de modo completo y sintéticamente.

RESPUESTA

En este caso se puede efectuar la reducción suprimiendo en las dos variables las dimensiones centrales media y mediocre.

Con ello la tipología quedaría transformada en la siguiente, en la que se indican dentro de cada tipo su denominación completa y sintética.

Conducta moral	*Práctica religiosa*	
	Alta	*Baja*
Buena	Práctica alta, conducta buena. Practicantes ejemplares.	Práctica baja, conducta buena. Hombres honestos.
Mala	Práctica alta, conducta mala. Fariseos.	Práctica baja, conducta mala. Libertinos.

163

En una investigación sociológica norteamericana sobre el gobierno familiar de los hijos se establecen los siguientes tipos de ejercicio de la autoridad por los padres:

1. Autocrático. Mis padres me dicen exactamente lo que debo hacer.

2. Autoritario. Me oyen, pero adoptan sus decisiones por sí mismos.

3. Democrático. Tengo considerable oportunidad para adoptar mis propias decisiones, pero ellos tienen la palabra final.

4. Igualitario. Mis opiniones son tan importantes como las de mis padres para decidir lo que deberé hacer.

5. Permisivo. Puedo adoptar mis propias decisiones, pero mis padres desearían que tuviera en cuenta su opinión.

6. Dejar hacer. Puedo hacer lo que quiero, aun sin consideración a lo que piensan.

7. Ignorancia. Mis padres no se preocupan de lo que hago.

Realizar la sustrucción de todos estos tipos, buscando las variables, y sus dimensiones, que se podrían distinguir en ellos y formando una tipología.

RESPUESTA

Los elementos o aspectos principales que se barajan en estos tipos son, por un lado, la facultad de decisión y, por otro, la facultad de opinión. Estas pueden corresponder en dichos tipos a los padres, a los hijos, o a ambos a la vez. Por tanto, se puede construir la tipología pedida considerando como variables la "facultad de decisión" y la "facultad de opinión", y como dimensiones de cada variable: padres, ambos. hijos.

Tendremos, pues:

FACULTAD DE DECISION	FACULTAD DE OPINION		
	Padres	*Ambos*	*Hijos*
Padres	1. Facultad de decisión y de opinión, los padres. *Autocrático.*	2. Facultad de decisión, padres. Facultad de opinión, ambos. *Autoritario.*	3. Facultad de decisión, padres. Facultad de opinión, hijos. *Democrático.*
Ambos	4. Facultad de decisión, ambos. Facultad de opinión, padres.	5. Facultad de decisión y opinión, ambos. *Igualitario.*	6. Facultad de decisión, ambos. Facultad de opinión, hijos.
Hijos	7. Facultad de opinión, padres. Facultad de decisión, hijos. *Permisivo.*	8. Facultad de decisión, hijos. Facultad de opinión, ambos. *Dejar hacer.*	9. Facultad de decisión y de opinión, hijos. *Ignorancia.*

164

Dados los siguientes tipos referentes a la posición de los obreros respecto de la empresa y los sindicatos, se pide realizar la sustracción de estos tipos, buscando las variables y sus dimensiones que se podrían distinguir en ellos y formando una tipología.

1. Dependencia o integración. Aceptación de la autoridad del patrón y apartamiento de la acción sindical.

2. Anárquico o marginal. Apartamiento de la acción sindical y rebelión frente a la empresa.

3. Revolucionario. Acción sindical y rebelión frente a la empresa.

4. Moderado. Acción sindical e integración en la empresa.

RESPUESTA

El examen de los tipos muestra que las variables que se conjugan en ellos son la aceptación de una autoridad de la empresa y la acción sindical. En ambas se pueden distinguir una dimensión positiva y otra negativa. Según esto se tiene la siguiente tipología:

ACCION SINDICAL	ACEPTACION DE LA AUTORIDAD DE LA EMPRESA	
	Sí (+)	No (—)
Sí (+)	+ + Moderados.	+ — Revolucionarios.
No (—)	— + Dependientes o integrados.	— — Anárquicos y marginales.

165

Dada la siguiente tabla tomada del estudio de Walis y Robert "How read a table" (1971, pág. 418), se pide efectuar su lectura y descripción.

TASAS DE ALFABETIZACION POR EDAD, SEXO Y COLOR. 1952.

Basada en una muestra de unas 25.000 personas. Fueron calificadas de analfabetas las personas incapaces de leer y escribir en cualquier lengua, con la salvedad de que la alfabetización fue supuesta para todos los que habían completado seis años o más de escuela. Sólo se incluye la población civil, no institucional, de catorce años de edad o más.

PORCENTAJES DE ALFABETIZACION

Edad/años	*Blancos*			*No blancos*			*Ambos colores*		
	Hombres	*Mujeres*	*Total*	*Hombres*	*Mujeres*	*Total*	*Hombres*	*Mujeres*	*Total*
14 a 24	1,2	0,5	0,8	7,2	1,4	3,9	1,8	0,6	1,2
25 a 34	0,8	0,6	0,7	9,7	3,8	6,4	1,6	0,9	1,2
35 a 44	1,2	0,5	0,8	7,5	5,9	6,6	1,7	1,0	1,3
45 a 54	2,2	1,4	1,8	12,8	10,4	11,5	3,2	2,3	2,7
55 a 64	3,6	3,4	3,5	19,4	16,9	18,1	4,7	4,4	4,5
65 en adelante	5,6	4,4	5,0	35,8	31,2	33,3	7,6	6,2	6,9
14 en adelante	2,1	1,5	1,8	12,7	8,2	10,2	3,0	2,1	2,5

Fuente: Statistical Abstract: 1955. Tabla 132, pág. 115. Fuente original, Bureau of the Census, Current Population Reports. Series p-20 No. 45.

RESPUESTA

Se aplican los siguientes pasos, según el estudio citado:

1. Leer el título cuidadosamente.
2. Leer atentamente las notas a la cabeza de la tabla o cualquier otra explicación.
3. Fijarse en la fuente o el origen de donde se ha tomado la tabla.
4. Fijarse en las notas de pie de página, si existen.
5. Hallar qué unidades son usadas en la tabla.
6. Fijarse en los promedios totales.
7. Ver la variabilidad que existe entre los datos de la tabla.
8. Mirar cómo el promedio está asociado con cada uno de los criterios principales de clasificación.
9. Examinar la consistencia de los efectos totales y la interacción entre ellos.
10. Finalmente, prestar atención a las aberraciones, anormalidades o irregularidades que se observen en los resultados de la tabla.

1) El título nos dice en este caso que la tabla refleja cuatro variables, una dependiente, las tasas de analfabetismo, y tres independientes, edad, sexo y color o raza, y que los datos se refieren al año 1952.

2) La nota a la cabeza de la tabla nos dice que la muestra es amplia, 25.000 personas, la noción adoptada de analfabetismo, y que sólo se tomó en cuenta el analfabetismo de los catorce años y más.

3) La lectura de la fuente muestra que es doblemente indirecta, pues se ha tomado de una revista, y ésta, a su vez, de otras fuentes de la Oficina de Estadística de Estados Unidos. Si bien ambas fuentes son de garantía.

4) No existen notas a pie de página.

5) Las unidades a que se refieren los números de las tablas son tasas porcentuales que expresan el número de analfabetos por cien habitantes.

6) Los promedios totales de la tabla se hallan en la columna marginal derecha para cada grupo de edad, y es para todos los grupos de edad de catorce años en adelante, 2,50 por 100, que es el promedio de la tasa general de analfabetismo derivada de la muestra en cuestión.

7) Ver qué variabilidad existe en los promedios. Es muy elevada, pues varía de 35,8 para los varones de sesenta y cinco años y más a 0,5 para las mujeres de catorce a veinticuatro años.

8) Mirar cómo los promedios están relacionados con cada uno de los criterios de clasificación, edad, sexo y color, o variables independientes.

EDAD.— En la edad se observa, atendiendo al total general, la columna extrema de la derecha, que la tasa de analfabetismo es aproximadamente la misma para los tres grupos de edad de catorce a cuarenta y cuatro años, pero a partir de aquí aumenta progresivamente hasta quintuplicar la cifra inicial en el grupo de sesenta y cinco años y más.

SEXO.— La comparación de las columnas de hombres y mujeres para ambas razas indica que en todas ellas la tasa de analfabetismo femenina es sensiblemente inferior, sobre todo para algunos grupos de edad, como el de catorce a veinticuatro, si bien estas diferencias podían ser debidas a las diferencias de distribución en edades, lo que debería ser investigado.

RAZA.— Comparando los porcentajes de las columnas de blancos y no blancos, se ve que en todas ellas hay también una gran diferencia, que va, en sus límites extremos, de 5 en los blancos a 33,3 en los no blancos, de sesenta y cinco años o más en ambos.

9) Examen de la consistencia de los efectos generales y la interacción entre ellos. La comparación de los incrementos de analfabetismo asociados con la edad muestra que son mayores para los no blancos que para los blancos, y que son aproximadamente los mismos para hombres y mujeres. Esto equivale a decir que el exceso de analfabetismo de los no blancos respecto a los blancos es mayor en las clases de mayor edad, y que la diferencia entre los sexos no se relaciona sistemáticamente con la edad.

Respecto al sexo, el estudio detallado conduce a la conclusión de que el exceso de tasa de hombres sobre las mujeres es más alto para no blancos que para blancos. Esto supone que las diferencias entre los colores son mayores para hombres que para mujeres.

10) Finalmente, buscar las aberraciones, anormalidades e irregularidades. La que más destaca es la referente a la edad de veinticinco a treinta y cuatro años en los blancos varones, en la que la tasa de analfabetismo desciende notablemente en relación a la anterior, en lugar de aumentar, como en las mujeres y negros.

Una explicación podría ser que este grupo proporcionó el núcleo de soldados de la segunda guerra mundial. Esto conduce a preguntarse si el Ejército enseñó a muchos analfabetos, y esta enseñanza afectó mucho más a los blancos que a los negros.

166

Dada la siguiente tabla que figura en la pág. 51, de "Tres estudios para un sistema de indicadores sociales". Fundación Foessa, Euramérica.

POSESIONES DE CIERTOS BIENES SEGUN LA CLASE SOCIAL SUBJETIVA

		Clase social subjetiva				
Bienes que poseen	*Total*	*Alta*	*Media alta*	*Media baja*	*Trabajadora*	*Pobre*
Ninguno	11	—	2	4	10	30
Radio o transistor	83	100	94	90	83	65
Frigorífico	28	86	66	39	19	5
Televisor	33	91	70	46	25	5
Lavadora eléctrica	37	91	71	71	52	7
Automóvil	12	77	39	16	5	—
Moto	11	9	11	13	13	5
Tocadiscos	12	82	38	17	5	(1)
Bicicleta	15	5	10	11	18	17
Total	100 (2.455)	(22)	(309)	(512)	(1.248)	(323)

Fuente: Datos del Informe de Foessa reanalizados. Se prescinde de 41 casos que no quieren o no saben contestar a la pregunta de identificación de clase.

Se pide efectuar su lectura y descripción.

RESPUESTA

1) Título. Indica que la tabla refleja dos variables, una dependiente, posesión de ciertos bienes, y otra independiente, clase social subjetiva. No se señala la comunidad a que se refieren los datos ni el año, pero como en la fuente se dice que los datos proceden del informe Foessa, se puede deducir que se refiere a España y al año en que fue realizado este estudio.

3) La fuente es como se indica el Informe Foessa. Los datos son de carácter primario, derivados de una encuesta realizada.

2) y 4) No existen notas en cabeza. Al pie de la tabla se nos dice que no se han tenido en cuenta, del total de la muestra, 41 encuestados. En relación a dicho total su número no es tan elevado como para que se pueda suponer la posibilidad de su influencia en los resultados tabulados.

5) Las unidades consignadas en la tabla son tantos por ciento de la población considerada. El 1 entre paréntesis no indica porcentajes, sino que sólo hay 1 encuestado de la clase pobre que tiene tocadiscos.

6) Los marginales o porcentajes totales de la tabla indican que es relativamente elevado el porcentaje de los que no poseen ningún bien de los que comprende la tabla y que los más generalizados son la radio o transistor, y los menos automóvil, moto y tocadiscos, 12, 11 y 12 por 100, respectivamente.

7) La variabilidad de los porcentajes presenta un recorrido máximo. Va desde el porcentaje cero (sólo hay un individuo que lo posea) para el tocadiscos en la clase pobre, al 100 por 100 de posesión de radio en la clase alta.

8) La variación de los porcentajes según los criterios de clasificación. Se observa que todas las clases, menos la trabajadora y la pobre, superan el porcentaje medio general en los distintos bienes, excepto en la bicicleta; respecto a este bien sólo superan la media general la clase trabajadora y la pobre. La clase trabajadora también supera la media general en la posesión de moto.

9) Consistencia. Excepto en estos dos bienes, el tanto por ciento de posesión de los demás desciende con el descenso de clase. Donde es más acusado el descenso en la posesión con la clase es en la posesión de automóvil y tocadiscos, que pasa del 77 por 100 en la clase alta, al 39 en la media alta, 16 en la media baja y 5 en la trabajadora y 0 en la pobre, respecto al tocadiscos pasa del 82 por 100 al 38, 17, 5 y 1 por 100.

10) Irregularidades. Las observaciones respecto a la moto y bicicleta se pueden considerar normales. Lo que llama más la atención son los tantos por cienco de la posesión de tocadiscos, inferiores al de los demás bienes por lo general, incluso automóvil, y que parece indicar un desinterés por la música y, por tanto, cultural, en relación a otros bienes de precio mucho más elevado, como el automóvil.

21. ANALISIS TIPOLOGICO O DE CLUSTERS

NATURALEZA, NOCION Y ELEMENTOS

Respecto a las técnicas de investigación social, interesa estudiar no sólo la clasificación como fase del proceso de investigación social, sino también las nuevas técnicas clasificatorias conocidas con el nombre de análisis de *clusters* o tipológico, que tienen como finalidad la formación de clases, tipos o grupos partiendo de un conjunto de datos. Esta es la razón de que se trate de ellas a continuación de la fase de clasificación, aunque, por su naturaleza, también participen de la condición de técnicas de análisis multivariable de datos.

La naturaleza del análisis tipológico viene determinada por las tres propiedades siguientes de ser una técnica

1. De análisis multivariable;
2. De análisis de datos, y
3. De tipo clasificatorio.

Por ser una técnica de análisis multivariable, no está limitada a datos sobre una o dos variables. Por ser una técnica de las llamadas de "análisis de datos" presenta un carácter inventivo, heurístico, inductivo y exploratorio, frente al carácter decisionista, deductivo y confirmatorio de las técnicas de análisis multivariable de tipo "clásico".

Por ser de tipo clasificatorio, se orienta decididamente a establecer grupos, tipos o clases. El análisis de clusters tiene por objeto, de acuerdo con Brian Everitt (1977, 1), idear, dada una muestra de N objetos o individuos, medidos cada uno de ellos respecto a p variables, un esquema clasificatorio para agrupar los objetos en n clases. El esquema indicado no responde a ninguna idea preconcebida. Antes bien, de lo que se trata es de descubrir la agrupación natural de los items o variables.

Según la última noción, los elementos fundamentales que intervienen en el análisis de clusters son:

a) El conjunto inicial a clasificar;

b) Los elementos que forman este conjunto;

c) Las variables medidas respecto a cada elemento;

d) Los grupos resultados del análisis.

El grupo inicial a clasificar está formado por un conjunto *N* de individuos. Los elementos del conjunto a clasificar pueden ser de todo tipo: hombres, comarcas, pueblos, ciudades, animales, plantas, etc. Lo mismo sucede con las variables, las cuales pueden ser, según su escala, nominales, ordinales, de intervalo y de razón.

El resultado del análisis puede consistir, o bien en la partición del conjunto inicial en una serie de subconjuntos, o bien en una ordenación jerárquica de los subconjuntos o subgrupos obtenidos. Su representación gráfica se puede realizar, como veremos, mediante, dendogramas o árboles.

TIPOS

Se suelen distinguir (Zaltmann y Burger) los dos tipos básicos siguientes:

El primero procede dividiendo la población total sucesivamente en subgrupos significativos.

En el segundo, en cambio, se van agrupando sucesivamente unos individuos a otros según las semejanzas que presentan.

Las técnicas de análisis del primer tipo reciben el nombre de divisivas, de segmentación o fraccionamiento. En ellas se sigue un proceso descendente, mediante la realización de dicotomías sucesivas en la población inicial.

Por el contrario, las técnicas del segundo tipo reciben el nombre de aglomerativas y también el de análisis factorial Q, para distinguirlo del análisis factorial clásico referente a variables, que se designa R. En estas técnicas, se sigue un proceso ascendente. Las clases se forman por aglomeraciones sucesivas de individuo a individuo o de individuo a grupo.

PROBLEMAS PREVIOS

Antes de exponer alguna de las técnicas más corrientes referentes a los tipos de análisis de clusters indicados, es preciso aludir a los dos problemas previos siguientes:

1. Medida de distancias, y
2. Medida de semejanza.

Medida de distancias

La medida de distancias o la de semejanzas es importante en el análisis de clusters porque muchas de sus técnicas, especialmente las aglomerativas, tienen como punto de partida el establecimiento de una matriz de distancias o de semejanzas. Las técnicas de medida de distancias son varias, pero quizás la más familiar y usada es la euclideana métrica. Su fórmula es la siguiente:

$$d_{ij} = \left[\Sigma \, (X_{ik} - X_{jk})^2 \right]^{1/2}$$

donde X_{ik} y X_{jk} son el valor de las variables i y j del elemento k.

El uso de esta técnica con base en cifras absolutas tiene el inconveniente de que puede dar resultados divergentes, si se cambia la escala de la variable utilizada, es decir, si, por ejemplo, en relación a la variable peso, en lugar de la escala kilogramos se usa la escala libras. Para obviar este inconveniente se recomienda normalizar los datos brutos iniciales, dividiéndolos por la desviación típica de la variable a la que se refieren, según esta fórmula:

$$Z_{ik} = (X_{ik})/\sigma_k$$

(Ver Ejercicios 167, 168)

Medidas de semejanza

Las medidas de semejanza son coeficientes de correlación pero referidos a individuos y no a variables, como dichos coeficientes.

De acuerdo con B. Everitt (1977, 51) un coeficiente de semejanza mide la relación entre dos individuos respecto a los valores de ambos en un conjunto *p* de variables.

Como coeficientes de asociación que son, pueden variar de 0 a 1. En muchos casos en los que las variables son dicotómicas, con dos dimensiones, presencia-ausencia, 1-0, se disponen los datos en una tabla de 2 x 2.

Existen diversos coeficientes para hallar la similitud en este caso, siendo quizás los más usados los siguientes:

$$\frac{a}{a+b+c} \qquad \frac{2a}{a+b+c}$$

a	b
c	d

Las variables cualitativas que tengan más de dos categorías pueden ser tratadas como dicotómicas o binarias, en las que sucesivamente cada una de las categorías se contrapone a la suma de las restantes.

En el caso de variables cuantitativas, la medida de semejanza usada comúnmente es el coeficiente de correlación producto momento r de Pearson.

Conforme ponen de relieve Johnson y Wickern (1982, 540) es posible siempre, pero no al revés, pasar de las distancias al coeficiente de semejanza, mediante, por ejemplo, la siguiente fórmula:

$$S_{ik} = \frac{1}{1 + d_{ik}}$$

siendo s_{ik} el coeficiente de semejanza entre i y k y d_{ik} su distancia. (Ver Ejercicio 169)

TECNICAS AGLOMERATIVAS

Las diversas técnicas de análisis de clusters de tipo aglomerativo existentes presentan los siguientes rasgos comunes:

1. Comienzan con el cómputo de una matriz de distancias o de semejanzas, entre los individuos o cosas a clasificar.
2. A partir de esta matriz se realiza la fusión, con los más próximos o similares, de individuos con individuos o de individuos con grupos.
3. El resultado final se concreta en un dendograma, que muestra gráficamente la sucesiva unión de individuos y subgrupos hasta formar un grupo único.

La diferencia entre ellas se encuentra en la segunda etapa, en la forma de realizar la fusión o agrupación. Entre los procedimientos diversos que se pueden utilizar al efecto destacan los siguientes:

a) El procedimiento del vecino más próximo, en el que se van agrupando los individuos más próximos o que estén a menor distancia.

b) El procedimiento del vecino más lejano, en el que, al contrario, se agrupan sucesivamente los individuos que estén a mayor distancia.

La técnica a seguir en ambos casos se describe en el ejercicio 167 y 168.

TECNICAS DIVISIVAS

Proceden, como se ha indicado, de manera descendente, por divisiones sucesivas de una población. Se dividen en dos ramas: las monotéticas que se basan sucesivamente en la posesión o no de cada variable o atributo singular, y las politéticas, que se fundan en los valores poseídos por los sujetos en todas las variables.

Dentro de estas técnicas, cabe destacar el análisis de segmentaciones. Su objetivo es encontrar la división mejor de un conjunto en dos subconjuntos, según la presencia o ausencia de una variable, de modo que un subconjunto la posea y el otro no.

El procedimiento más común para lograr este objetivo en el caso de variables con solo dos categorías es hacer la división por el atributo o variable que haga máximo el ΣX^2_{jk}. En lugar de éste, también se puede usar la variable que haga máxima los sumatorios siguientes:

$$\text{Max } [\Sigma \, |ad - bc|] \text{ ó Max } [\Sigma \, (ad - bc)^2]$$

(Ver ejercicio 169)

EJERCICIOS

167

Dada la siguiente tabla:

Individuo	*Renta en cientos de miles de ptas.*	*Indice de prestigio social*
A	*8*	*10*
B	*12*	*15*
C	*6*	*9*
D	*20*	*17*
E	*35*	*25*

a) Hallar las distancias euclideanas métricas de los individuos de la tabla entre sí.

RESPUESTA

Aplicando la fórmula:

$$d_{ij} = \Sigma (X_{ik} - X_{jk})^2]^{1/2}$$

a cada una de las diez distancias posibles se tiene:

$$D^2_{AB} = (8 - 12)^2 + (10 - 15)^2 = 16 + 25 = 41$$

El exponente 1/2 de la fórmula queda compensado con la elevación al cuadrado de D.

$$D^2_{AC} = (8 - 6)^2 + (10 - 9)^2 = 4 + 1 = 5$$

$$D^2_{AD} = (8 - 20)^2 + (10 - 17)^2 = 144 + 49 = 193$$

$$D^2_{AE} = (8 - 35)^2 + (10 - 25)^2 = 729 + 225 = 954$$

$$D^2_{bC} = 72$$

$$D^2_{BD} = 68$$

$$D^2_{BE} = 629$$

$$D^2_{CE} = 1097$$

$$D^2_{DE} = 289$$

b) Construir la matriz de las distancias obtenidas.

RESPUESTA

Se trata de una matriz cuadrada del mismo número de filas y columnas, que corresponden al número de individuos y simétrica porque sus entradas se repiten simétricamente a izquierda y derecha de la matriz principal, ya que la distancia de A a B es igual que la de B a A, etc.

	A	B	C	D	E
A	0	41	5	193	964
B		0	72	68	629
C			0	260	1097
D				0	289
E					0

c) Empezar el análisis de cluster de esta matriz por el procedimiento del vecino más próximo, formando el primer subgrupo.

RESPUESTA

Es siempre el formado por los dos individuos a menor distancia en la matriz, que, en este caso, es cinco y corresponde a A y C.

d) Hallar las distancias del subgrupo AC a los demás individuos y formar la nueva matriz de distancias.

RESPUESTA

Aplicando la fórmula,

$d_{(ij)k}$ = min. (d_{ik}, d_{jk}), se tiene

$d_{(AC)B} = \min (d_{AB}, d_{CB}) = \min (41, 72) = 41$

$d_{(AC)D} = \min (193, 260) = 193$

$d_{(AC)E} = \min (964, 1077) = 964$

Nueva matriz de distancias:

	AC	B	D	E
AC	0	41	193	964
B		0	68	629
D			0	289
E				0

e) Formar el segundo subgrupo y hallar las nuevas distancia y matriz.

RESPUESTA

El subgrupo a formar es el que señale la menor entrada o valor de la matriz que es 41 y corresponde a (AC) y B, que se deben unir formando (ABC).

Las distancias, aplicando la fórmula de siempre serán:

$$d_{(ACB)E} = (964, 629) = 629$$

Matriz:

	(ACB)	D	E
(ACB)	0	68	629
D		0	289
E			0

f) Formar el tercer subgrupo y hallar las distancias y nueva matriz.

RESPUESTA

Es como siempre la menor distancia, 68, que corresponde a (AC.B) y D. Luego D debe pasar a engrosar el subgrupo de ACB, formando ACBD.

Como únicamente queda otro individuo en el conjunto, sólo puede haber una menor distancia, la existente entre ACBD y E, que pasará a formar parte del grupo final.

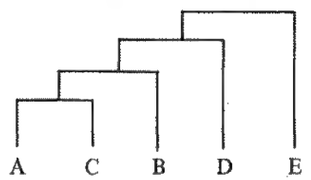
g) Formar el dendograma correspondiente a los resultados obtenidos en el análisis de cluster realizado.

RESPUESTA

El dendograma tiene que reflejar, mediante líneas de enlace, el proceso seguido por la agrupación sucesiva derivada del análisis de cluster realizado.

El proceso ha consistido en la unión primero de los elementos AC a los que se han ido añadiendo sucesivamente los individuos BDE.

Por tanto, el dendograma final deberá tener la forma siguiente:



168

Dada la matriz de distancias formada en el ejercicio anterior, n. 2, realizar su análisis de clusters por el procedimiento del vecino más alejado, y para ello:

a) Formar el primer subgrupo y calcular sus distancias a los restantes individuos del conjunto.

RESPUESTA

El primer subgrupo, como en el ejercicio anterior, es el formado por los individuos a menor distancia según la matriz, o sea, los individuos A y C.

Para hallar las distancias de este subgrupo a los restantes individuos se aplica la fórmula correspondiente, que se diferencia de la anterior por elegirse los máximos y no los mínimos de los pares de valores.

$$d_{(ij)k} = \max.\ (d_{ik} - d_{jk})$$

$$d_{(AC)\in} = \max\ (d_{AB} - d_{CB}) = \text{Max}\ (41,72) = 72$$

$$d(AC)D = \max\ (193, 260) = 260$$

$$d_{(AC)E} = \max\ (964, 1097) = 1097$$

b) Formar la nueva matriz de distancias.

RESPUESTA

Sería la siguiente:

	AC	B	D	E
AC	0	72	260	1097
B	72	0	68	269
D	360	68	0	289
E	1097	269	289	0

c) Formar el segundo subgrupo y hallar las nuevas distancias y matriz.

RESPUESTA

El nuevo subgrupo es el que corresponda a la menor entrada de la matriz que es 68, y pertenece a los individuos B y D.

Las nuevas distancias serán:

$$d_{(AC)\ (BD)} = \max\ (d_{(AC)B},\ d_{(AC)D}) \times \max\ (72, 260) = 260$$

$$d_{(AC)E} = 1097 \text{ como antes}$$

$$d_{(BD)E} = \text{Max}\ (68, 269) = 269.$$

Nueva matriz:

	AC	BD	E
AC	0	260	1097
BD	260	0	269
E	1097	269	0

d) Formar el tercer subgrupo y hallar las distancias y nueva matriz.

RESPUESTA

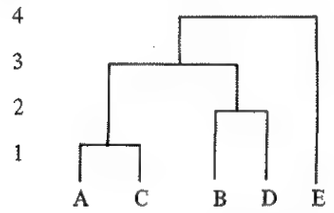
Es como siempre el de la menor distancia, 260, que corresponde a (BD) y (AC) por lo que BD pasa a formar parte del subgrupo AC.

Después de esto, sólo quedan dos subgrupos, DACBD y E, por tanto, con una sola distancia entre ellos, que siempre será la menor, lo que indica que ambos subgrupos deben unirse para formar el grupo final.

e) Formar el dendograma correspondiente.

RESPUESTA

El proceso seguido en la agrupación, que debe ser reflejado por el dendograma, es la unión, primero de A y C, después de B D, luego de BD con AC, y finalmente de AC y Nd con E. El dendograma por tanto será:



Se ve, por tanto, que el resultado final de esta técnica no siempre coincide con el de la anterior.

169-71

Dada la siguiente tabla, que expresa las categorías (1-0) de diez individuos en tres variables dicotómicas,

Individuos	*Variables*		
1	1	0	1
2	0	1	1
3	1	1	1
4	0	0	0
5	1	0	0
6	0	0	1
7	1	1	0
8	0	1	1
9	1	0	1
10	1	1	1

a) Iniciar su análisis de asociación o de segmentaciones, formando las tablas de contingencia que expresan las frecuencias que se den, respecto a las combinaciones posibles de las categorías de las dos variables en los individuos.

RESPUESTA

Hay que relacionar en las tablas de contingencia todas las variables dos a dos, contando las veces que se dan en la tabla inicial, las cuatro combinaciones posibles de las categorías de las dos variables: (1-1; 1-0; 0-1; 0-0).

Variable 1	*Variable 2*	
	1	0
1	3	3
0	2	2
	N	10

Variable 1	*Variable 3*	
	1	0
1	4	2
0	3	1
	N = 10	

Variable 2	*Variable 3*	
	1	0
1	4	1
0	3	2
	N – 10	

b) Hallar los χ^2 de todas estas tablas.

RESPUESTA

Se aplica la fórmula siguiente, especial para tablas de 2 x 2:

$$\chi^2_{jk} = \frac{(ad - bc)_2 \, N}{(a+b)(a+c)(b+d)+(c+d)}$$

Variables 1 y 2:

$$\chi^2_{11} = \frac{(6-6)^2\ 10}{(6)\,(5)\,(5)\,(4)} = 0$$

a	b
c	d

Variables 1 y 3

$$\chi^2_{13} = \frac{(4-6)^2\ 10}{(6)\,(7)\,(3)\,(4)} = \frac{.40}{504} = 0{,}079$$

Variables 2 y 3

$$\chi^2_{23} = \frac{(8-3)^2\ 10}{(7)\,(5)\,(5)\,(3)} = \frac{250}{525} = .47$$

c) Determinar la variable que ha de servir de criterio para dividir los sujetos en los dos primeros subgrupos.

RESPUESTA

Para hallarla, se suman dos a dos los χ^2 obtenidos. De la observación de las sumas obtenidas se deduce que la variable que hace máximo las sumas del χ^2 es la tercera, pues es la que está presente de forma repetida en la suma máxima obtenida.

d) Dividir el conjunto inicial en dos subgrupos: según el criterio de división hallado.

RESPUESTA

El criterio de división es la presencia o ausencia, es decir, los valores 1 ó 0 de la variable 3, en la tabla inicial, de tal modo que los individuos que tengan 1 en dicha variable pasarán a formar parte de un subgrupo y los que tengan 0 de otro. Por tanto, el primer subgrupo estará formado por los individuos 1, 2, 3, 6, 8, 9 y 10 y el segundo por los 4, 5 y 7.

e) Proseguir el análisis respecto a los individuos del primer subgrupo, formando la nueva tabla de partida y las tablas de contingencia.

RESPUESTA

Tabla de partida

Individuos	*Variables*		
	1	2	3
1	1	0	1
2	0	1	1
3	1	1	1
6	0	0	1
8	0	1	1
9	1	0	1
10	1	1	1

Tablas de contingencia

Variable 1	*Variable 2*	
	1	0
1	2	2
0	2	1

Variable 1	*Variable 3*	
	1	0
1	4	0
0	3	0

Variable 2	*Variable 3*	
	1	0
1	4	0
0	3	0

f) Hallar los χ^2 de estas tablas.

RESPUESTA

$$\chi^2_{12} = \frac{(4-2)^2\ (7)}{4\ (3)\ (4)\ (3)} = \frac{28}{144} = .19$$

$$\chi^2_{13} = 0; \quad \chi^2_{23} = 0$$

g) Determinar la variable que ha de servir de criterio para dividir los sujetos de este subgrupo.

RESPUESTA

Es la variable 1, pues la que hace máxima la suma de los χ^2 dos a dos.

h) Dividir el conjunto inicial en dos subgrupos según el criterio de división hallado.

RESPUESTA

Es la presencia o ausencia, 1 ó 0, de la variable 1 en la tabla de partida.

Por tanto, el primer subgrupo será formado por los individuos 1, 3, 9 y 10 y el segundo por el 2, 6 y 8.

Por esto, los subgrupos formados por los individuos más similares entre sí, respecto de las variables investigadas, resultantes de la 1ª y 2ª división realizada son tres:

4, 5 y 7; 2, 6 y 8; 1, 3, 9 y 10

Sección 6
ANALISIS

f) Hallar los χ^2 de estas tablas.

RESPUESTA

$$\chi^2_{12} = \frac{(4-2)^2\ (7)}{4\ (3)\ (4)\ (3)} = \frac{28}{144} = .19$$

$$\chi^2_{13} = 0; \quad \chi^2_{23} = 0$$

g) Determinar la variable que ha de servir de criterio para dividir los sujetos de este subgrupo.

RESPUESTA

Es la variable 1, pues la que hace máxima la suma de los χ^2 dos a dos.

h) Dividir el conjunto inicial èn dos subgrupos según el criterio de división hallado.

RESPUESTA

Es la presencia o ausencia, 1 ó 0, de la variable 1 en la tabla de partida.

Por tanto, el primer subgrupo será formado por los individuos 1, 3, 9 y 10 y el segundo por el 2, 6 y 8.

Por esto, los subgrupos formados por los individuos más similares entre sí, respecto de las variables investigadas, resultantes de la 1ª y 2ª división realizada son tres:

4, 5 y 7; 2, 6 y 8; 1, 3, 9 y 10

Sección 6
ANALISIS

22 . ANALISIS, INTERPRETACION Y EXPLICACION

NOCION DEL ANALISIS, LA INTERPRETACION Y LA EXPLICACION SOCIOLOGICOS

La clasificación presenta los resultados del estudio realizado, agrupados y relacionados de acuerdo con los objetivos de la investigación. Representa, por tanto, una síntesis de los hechos observados, no única, pues los datos son susceptibles de múltiples clasificaciones. Pero aunque en la clasificación se contienen, ordenadas y dispuestas de la forma que se ha creído conveniente, las informaciones recogidas en el trabajo de campo, no acaba en ella el proceso investigador.

A continuación es necesario deducir las notas y conclusiones sociológicas que se derivan de toda la masa de datos clasificada y ordenada. Las tablas, por tanto, precisan ser analizadas, es decir, estudiadas en sus distintos elementos y aspectos, a fin de establecer las consecuencias que se pueden deducir de ellas respecto a la finalidad de la investigación y a las hipótesis que se han pretendido contrastar.

Se trata de una operación importante que remata y culmina todo el proceso de investigación y en la que se recogen y acumulan sus frutos y resultados.

Pero no es una operación simple sino compleja, pues se pueden distinguir en ella tres aspectos principales, el análisis propiamente dicho, la interpretación y la explicación.

El análisis y la interpretación tienen un alcance específico, en cuanto se trata en ellos ante todo del estudio de los datos obtenidos mediante la observación y clasificación. Como se ha indicado, el análisis busca hacer explícitas las propiedades, notas y rasgos de todo tipo que con relación a las variables estudiadas, se derivan de las tablas en las que se condensa la clasificación, mientras que en la interpretación se trataría de determinar la significación y alcance sociológicos de dichas propiedades y rasgos.

La explicación tiene un sentido más general. Respecto a la investigación, se puede considerar como la expresión teórica última y general de los resultados de la investigación o, si se quiere como el esquema lógico general de la solución al problema planteado que aquéllos representan.

ANALISIS E INTERPRETACION Y TEORIZACION

El análisis, según lo anterior, se basa inmediatamente en las tablas numéricas. De ahí se deriva el carácter inicialmente cuantitativo que presenta.

Da lugar a la formulación, en la terminología de J. A. Davis (1971), de *findings*. Según este autor *findings* son afirmaciones sobre las propiedades estadísticas de los datos obtenidos en las operaciones de investigación.

Un ejemplo de finding que da Davis es el siguiente: "los individuos con bajas puntuaciones en los tests de inteligencia tienden a presentar puntuaciones altas en las escalas de prejuicio".

La interpretación ofrece, por el contrario, un carácter cualitativo más acusado. Su punto de partida son los *findings*, las afirmaciones sobre las propiedades estadísticas de los resultados.

Con base en ellos la interpretación establece conclusiones que son enunciados sobre los términos conceptuales de las variables.

La conclusión que se puede inferir del ejemplo de *finding* expuesto sería la siguiente: Baja inteligencia conduce a hostilidad hacia los grupos sociales extraños.

Se puede afirmar que tanto el análisis como la interpretación, son las fases fundamentales del proceso de teorización, que, según se expuso en el capítulo tercero, constituye la segunda gran etapa de la investigación.

En este proceso el análisis transforma los datos numéricos de las tablas en enunciados de carácter estadístico como el indicado. Pueden ser referentes a una sola variable o a dos o más.

Por su parte la interpretación es la que realiza la teorización de los enunciados estadísticos.

En esta tarea se puede distinguir tres partes:

La primera es la transformación de los enunciados estadísticos sobre una variable en enunciados conceptuales teóricos.

La segunda es la conversión de los enunciados estadísticos sobre relaciones entre variables en proposiciones o enunciados teóricos referentes a relaciones entre variables y su contraste con las hipótesis formuladas.

La tercera es el encuadramiento de estas proposiciones en la teoría científica ya existente.

Este nuevo tratamiento tiene naturaleza principalmente lógica, de raciocinio. Su fin es completar el ciclo de la investigación, que se inició partiendo de la teoría, y hacer que los resultados confluyan nuevamente en la teoría. El objetivo de este tramiento es convertir los resultados numéricos de la investigación en enunciados teóricos y nuevas teorías.

Lo anterior supone un proceso de abstracción o generalización. Como escribe W. L. Wallace (1976, 56) debe considerarse que "las teorías surgen al hacerse más abstractos los términos y relaciones de las generalizaciones empíricas y también al introducir otros términos abstractos que se refieren a constructos no observables".

B. G. Glaser y A. L. Strauss (1974) defienden una teoría fundada en la investigación social misma, mediante la generalización de sus resultados a partir de los datos y explican en dicha obra una técnica al efecto a la que dan el nombre de "grounded theory", que se puede considerar como una ilustración de este proceso de teorización.

CLASES DE ANALISIS

Las formas de clasificación de los distintos tipos de análisis pueden ser muy diversas.

Aquí interesa estudiar las siguientes:

Por el número de variables, se tienen los análisis uni-, bi- y multi-variable, según se refiera a tablas de una, dos o más de dos variables, respectivamente.

De acuerdo con su carácter se puede hablar de los análisis exploratorios y confirmatorios, según pretendan, respectivamente hallar nuevas hipótesis o verificarlas.

Conforme a su cometido, se puede hablar de análisis funcional, causal, del cambio, de las decisiones, de la estructura de grupos pequeños, de la movilidad social, según pretendan el estudio de las causas, el cambio, las decisiones, etc.

Por último, si se tiene en cuenta su naturaleza, se pueden distinguir: análisis cuantitativos, estadísticos o matemáticos, de los análisis cualitativos de carácter lingüístico o de otro tipo.

TIPOS DE ANALISIS SEGUN SU	Núm. de variables	Univariable Bivariable Multivariable
	Carácter	Exploratorio Confirmatorio
	Naturaleza	Cuantitativo Cualitativo
	Objetivo	Causal Del cambio De decisiones De grupos De sistemas

ANALISIS Y TIPOS DE DATOS

Los datos objeto de análisis pueden haber sido obtenidos mediante nuestro propio trabajo de investigación o pueden proceder de otras investigaciones o fuentes. Los primeros reciben el nombre de datos primarios y, los segundos, de datos secundarios.

Asímismo, en ambos casos puede tratarse de datos no numéricos de carácter cualitativo o pueden tener carácter numérico o cuantitativo.

Se tienen, de este modo, las siguientes categorías de datos:

1. Datos primarios numéricos.
2. Datos primarios no numéricos.
3. Datos secundarios numéricos, y
4. Datos secundarios no numéricos.

Los primeros se pueden caracterizar por su elevado coste de obtención; su mayor adecuación normalmente al objeto de investigación; el mayor conocimiento y control de la fiabilidad que merecen y la dificultad de obtener series temporales de datos de este tipo.

Los primarios no numéricos se consideran, en general, difíciles de obtener, dada la frecuente dispersión de sus fuentes; suelen exigir análisis muy técnicos; el control de la fiabilidad que merecen es escaso y es también muy difícil obtener series temporales de ellos. Su obtención puede resultar más sesgada, dado su carácter cualitativo, por la

ideología del investigador. Su utilidad en la investigación suele tener caracter exploratorio y servir para sugerir hipótesis nuevas o apoyar hipótesis previas.

Por último, los datos secundarios numéricos y no numéricos se caracterizan por el menor coste económico de su obtención con relación a los primarios, aunque pueden a veces exigir bastante trabajo.

En general, en ellos son menos conocidos sus coeficientes de error y existe menos control de su fiabilidad, que en los primarios.

En los datos secundarios cuantitativos suele ser menos difícil la obtención de series temporales.

ANALISIS E INTERPRETACION UNIVARIABLE

El análisis estadístico de los datos referentes a una sola variable o de las distribuciones de frecuencias proporciona al analista, como es sabido, fundamentalmente:

a) Medidas representativas de la distribución;

b) Indices de la dispersión o desviación típica de estas medidas respecto a la distribución que representan,

c) Procedimientos para normalizar los valores de la distribución; y

d) Medidas de las desigualdades de unos valores en relación con otros.

Con base en estas medidas el análisis estadístico puede establecer enunciados que expresen sus resultados.

Después, la interpretación, mediante la comparación de los resultados obtenidos entre sí o con otros referentes a lugares y tiempos distintos, así como mediante el enjuiciamiento de los mismos a través de la teoría existente respecto a la materia en cuestión, deberá determinar su importancia y significación teóricas y formular en términos científicos los resultados.

En consecuencia, el análisis e interpretación de los resultados sobre una sola variable, podría comprender los siguientes pasos:

1. Se parte de los resultados estadísticos.
2. Es conveniente su normalización para facilitar comparaciones.
3. Comparación interna de unos valores o categorías de la variable analizada con otros, y de las medidas estadísticas con las referentes a otras poblaciones.
4. Expresión de enunciados estadísticos que reflejan las conclusiones derivadas del análisis estadístico.

5. Su interpretación según el marco o las condiciones y supuestos de la investigación.
6. Dar una formulación teórica a estos enunciados.
7. Contraste y armonización con la teoría ya existente sobre la materia. (Ver ejercicio 172).

ANALISIS E INTERPRETACION BIVARIABLE

De modo similar al caso univariable, el análisis estadístico proporciona:

a) Coeficientes que señalan el grado y la forma de asociación entre las dos variables; y

b) Mediante los test estadísticos de hipótesis indica la significatividad de estos coeficientes con relación a la población de donde proceden las muestras.

Esta determinación puede dar lugar, de modo similar al caso anterior, a la formulación de enunciados estadísticos que expresen inmediatamente su importancia cuantitativa, y servir de base a la interpretación de éstos, mediante su comparación con otros referentes a tiempos y lugares distintos y el enjuiciamiento de los mismos a través de la teoría existente respecto a la materia en cuestión.

A modo de síntesis y como simple orientación, en el análisis e interpretación de los datos sobre dos variables, se pueden distinguir los siguientes pasos:

a) El análisis estadístico debe comenzar con la lectura y estudio detenido de la tabla, efectuando comparaciones dentro y entre los valores o frecuencias de las diversas categorías de las variables que las forman, y seguir con la determinación del coeficiente de asociación, junto con el análisis de regresión siempre que se pueda efectuar.

b) Los coeficientes de asociación no necesitan ser normalizados para hacer comparaciones con los de otras muestras o poblaciones, al contrario de los coeficientes de regresión que sí deben ser normalizados con el fin indicado.

c) En el caso de datos de muestra se deberá hallar el error muestral y contrastar la significatividad de los coeficientes hallados y especialmente si son distintos de cero, mediante el test de hipótesis adecuado.

d) Es aconsejable realizar alguna forma de representación gráfica.

e) Se estudiará el tipo de relación que exista entre las variables, simétrica, asimétrica, recíproca.

f) Se expresarán los enunciados que reflejan las conclusiones que se deriven del análisis estadístico.

g) Se dará una formalización teórica a estos enunciados.

h) Su interpretación según el marco o las condiciones y supuestos de la investigación.

i) Por último, se contratará y armonizarán los resultados teóricos obtenidos con la teoría ya existente. (Ver ejercicio 173).

ANALISIS E INTERPRETACION MULTIVARIABLE

El análisis bivariable cumple las funciones indicadas, pero, sobre todo en las ciencias sociales, es fundamentalmente incompleto porque en los fenómenos sociales las variables no se presentan simplemente en parejas sino en conjuntos inseparables.

De ahí que el hecho de separar dos variables de las demás a efectos de sus análisis sea una abstracción que no se ajusta a la realidad.

Sin negar que dicho análisis tenga un valor, sobre todo descriptivo, no descubre, ni profundiza, ni explica totalmente el carácter de la relación entre las dos variables, dado que dicha relación normalmente se ha de encontrar necesariamente afectada por la influencia de las restantes variables que intervienen en el fenómeno estudiado.

Por ello, el análisis y la interpretación multivariable, que tiene en cuenta no dos variables solamente sino tres o más, o en el que la relación entre dos variables se va probando mediante su puesta en relación con otras variables, se acerca más a la realidad y puede permitir una comprensión más exacta del fenómeno estudiado. (Ver ejercicios 174, 175).

NOCIONES DE EXPLICACION: LINGUISTICA, LOGICA, GNOSEOLOGICA, EPISTEMOLOGICA y ONTOLOGICA DE LA EXPLICACION

Desde un punto de vista científico se encuentran en los diversos autores de filosofía de la ciencia múltiples y diversas nociones o definiciones de la explicación.

En mi opinión, todas ellas, por lo general, en lugar de ofrecer concepciones divergentes de la explicación, lo que hacen es mostrar su sentido o significación desde diversos enfoques, niveles o puntos de vista, los cuales se pueden reducir a los cinco siguientes: lingüístico, lógico, epis-

temológico, gnoseologico y ontológico a los que se podría añadir el psicológico.

– En un plano lingüístico, la explicación es simplemente la expresión de un enunciado mediante otras palabras distintas pero más comprensibles.

Una explicación semántica, como escribe Kaplan (1964, 327) consiste en "un conjunto de palabras que tienen un significado similar y equivalente a aquéllas a ser explicadas, pero de más fácil comprensión".

Desde este punto de vista, la explicación es siempre dar una respuesta al interrogante, que explícita o implícitamente late en todo enunciado explicado, mediante otros enunciados de palabras más claras en su sentido.

– Enfocada lógicamente, la explicación es, como afirma Bunge (1975, 102) "una subsunción bajo un conjunto de premisas". Esta subsunción se debe dar, también implícita o explícitamente, en toda explicación, en cuanto ésta implica siempre la reducción de lo explicado a lo que explica, lo que no se puede realizar lógicamente de modo válido, sin premisas implícitas que avalen la equivalencia, aunque sólo sea de significado, entre los términos explicados y los que explican.

– Desde un punto de vista gnoseológico, la definición de explicación debe hacer referencia directa al conocimiento. De ahí que en este sentido "las explicaciones, según Nagel (1974, 109), puedan ser consideradas como intentos de comprender lo no conocido, en términos de lo conocido a algo más conocido".

Este punto de vista marca, a diferencia de los dos precedentes que señalan la forma lingüística y lógica de la explicación, el fin de toda explicación, que no puede ser otro sino el de hacer comprensible o más comprensible algo.

– Enfocada la explicación epistemológicamente, es decir, con referencia a la ciencia en cuanto conocimiento metódico de la realidad observable, las ciencias pretenden dar cuenta de hechos o clases de hechos estableciendo sus relaciones con otros hechos o clases de hechos, que los condicionan o determinan. "Las ciencias, afirma Nagel (1974, 17) tratan de descubrir y formular en términos generales las condiciones en las cuales ocurren sucesos de diverso tipo, y las explicaciones son los enunciados de tales condiciones determinantes. Sólo es posible lograr este objetivo distinguiendo o aislando ciertas propiedades en el tema estudiado y discerniendo los esquemas de dependencia reiterados que vinculan estas propiedades unas a otras".

– En un plano ontológico, la explicación consiste, en último término en determinar la causa de lo explicado. En este sentido, una explicación

como señala Rudner (1973, 96) "es una respuesta a la pregunta" ¿por qué?". Ahora bien, a esta pregunta se puede contestar en la explicación refiriéndose ya al origen, a lo que ha producido o dado lugar al fenómeno a explicar, o bien a la intención o fin pretendida por el agente productor, cuando éste es racional, o finalmente, determinando porqué es como es el fenómeno en cuestión, es decir, en términos modernos, qué variables lo forman y cuáles son sus relaciones o estructura.

– Finalmente, psicológicamente, la explicación se pide por alguien y se da para alguien. Desde este punto de vista, una explicación será tal si satisface las dudas de ese alguien, si le pone de manifiesto lo que desconocía y quería saber sobre alguna cuestión. En este caso se debe dar una conjunción entre el tipo de explicación y el nivel de la duda. Que satisfaga o no psicológicamente una explicación depende pues no solo de que sea correcta, sino de que se sitúe al mismo nivel que la duda a la que se pretende responder.

TIPOS DE EXPLICACION

Las explicaciones se pueden clasificar según,

a) Su objeto,

b) Su forma lógica y alcance real, y

c) Su nivel científico.

a) Teniendo en cuenta su objeto, las explicaciones pueden serlo:

1. De conceptos y enunciados,
2. De un acontecimiento o suceso,
3. De clases de acontecimientos o sucesos,
4. De hechos, en el sentido de realidades ya existentes,
5. De acciones.

1. Las explicaciones de conceptos o enunciados no pueden ser otra cosa que explicaciones de tipo lingüístico o semántico, mediante las cuales expresamos su sentido o significación con otras palabras que se suponen más claras. Según esto, serían explicaciones de este tipo las definiciones.

2. Los acontecimientos o sucesos son, en general, hechos o realidades que, en lugar de poseer como los demás un estado o existencia previo determinado, más o menos permanente, se caracterizan por referirse

específicamente a lo que ocurre o acaece en un momento determinado. Su explicación se refiere preferentemente no a la realidad en sí que representa lo acontecido, sino a lo que ha dado lugar a que se produjera en ese momento el acontecimiento en cuestión, v.g., que un coche se salga de la carretera y choque contra un árbol, aunque también pueden entrar en esta explicación los efectos producidos.

Las explicaciones típicas de un acontecimiento o suceso dan cuenta de éste, según señala Harre (1967, 34), informando sobre otro acontecimiento. Por ejemplo: El fuego se ha apagado (un suceso) porque la chimenea no tira (otro suceso). Harre llama a estas explicaciones lineales porque lo explicado, *explicandum*, y lo que explica, *explicans*, presentan el mismo nivel lógico, en cuanto el *explicandum* y el *explicans* se mantienen en lo singular.

Son explicaciones vulgares no científicas.

3. Las explicaciones de clases de acontecimientos o sucesos son las que valen no sólo para un acontecimiento o suceso singular, sino para todo un conjunto de ellos. Es, siguiendo con el ejemplo de Harre, como si explicaramos el suceso de que este fuego se haya apagado diciendo que "los fuegos se apagan porque las chimeneas se obturan". Aquí, la explicación es de tipo general, una ley, y por ello su nivel lógico es distinto del hecho singular de haberse apagado este fuego que se intenta explicar. Se basan, pues, como las explicaciones científicas, en la formulación de leyes.

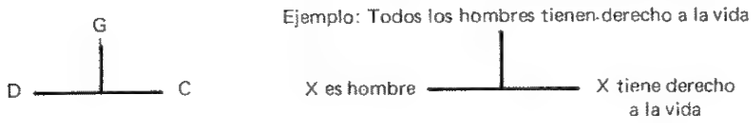
4. Las realidades existentes se pueden explicar por los agentes o causas que las han producido, por los elementos que las forman y la estructura de sus relaciones, por las funciones de sus elementos o de toda la realidad en un conjunto más amplio. No se trata sino, respectivamente, de las explicaciones causales, estructurales y funcionales que se estudian a continuación.

5. Las explicaciones de acciones son también explicaciones de acontecimientos, pero con la particularidad de que en ellas interviene la libertad humana, las intenciones, propósitos y motivos del hombre. En ellas, por tanto, el carácter que se podría llamar "mecánico" de las explicaciones de sucesos puras, resulta modificado o alterado por la imprevisibilidad propia de la libertad del hombre. Estas explicaciones son las peculiares de las ciencias humanas y sociales.

b) Según el criterio de su forma lógica, se suelen distinguir cinco tipos de explicación: deductiva, probabilística, teológica, genética y estructural.

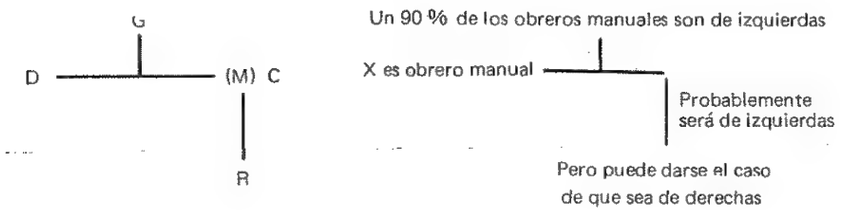
Cuando de las premisas se deduce necesariamente la conclusión se tiene la explicación deductiva.

Según Leónidas Hegenberg (1969) el esquema general básico de la explicación deductiva comprende una garantía, G, que es la regla general que se utiliza en la explicación; datos, D, o circunstancias de hecho relativas al caso considerado y una conclusión, C, que engloba el caso considerado en la regla general.



En las ciencias en general y sobre todo en las ciencias sociales, muchos de los fenómenos explicados no están contenidos necesariamente en la garantía, en la regla general, sino sólo probablemente, por eso, las explicaciones en estos casos se llaman explicaciones probabilísticas.

En este caso, la conclusión está afectada en el esquema de la explicación por un cualificador modal, M, que fundamenta la posibilidad que el caso examinado sea una excepción de la regla general y la conclusión por tanto no sea cierta respecto del mismo.



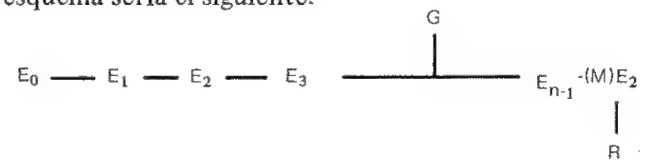
Explicaciones teleológicas funcionales o finalísticas son las que explican un acontecimiento, factor o fenómeno por las acciones que realiza en un todo para mantener su funcionamiento o por la finalidad que se pretende con una acción determinada.

La explicación teleológica es del tipo siguiente:

La función P, en un sistema S, con una organización H, es la de permitir, en la circunstancia C, la actividad A.

Las explicaciones genéticas consisten en establecer que un fenómeno es una consecuencia de un proceso o de una serie de acontecimientos previos.

Su esquema sería el siguiente:



E_n es el estado de cosas que requiere explicación; E_o es el estado inicial; E_i son los estadios intermedios.

En este tipo de explicación se tienen enunciados que afirman que sucedió el acontecimiento E_1 y enunciados generales estadísticos que afirman una dependencia entre diversas características de los E_i.

Estas explicaciones son muy débiles, ya que no consisten en último en otra cosa que en una secuencia de explicaciones probabilísticas.

Por último, las estructurales consisten en establecer cómo está estructurado un fenómeno, es decir, qué elementos lo componen y cuáles son sus relaciones, o en determinar la posición de un elemento en el conjunto de la estructura a la que pertenece.

c) Según su nivel científico, las explicaciones se pueden clasificar en científicas y no científicas, según utilicen o no, conforme se ha indicado antes, leyes generales y, además, en sentido estricto, según que la respuesta al porqué de la cuestión a explicar sea o no el resultado de una investigación científica. Un tipo de explicaciones que se pueden considerar como intermedias a las científicas y no científicas son las explicaciones o interpretaciones *"postfactum"*. Según Merton (1970, 103) ocurren con frecuencia en la investigación social empírica. Consisten en que "se recogen los datos y sólo más tarde se someten a comentario interpretativo". Se caracterizan, pues, por la formulación de una "interpretación después de haber hecho las observaciones" y no por la "comprobación empírica de una hipótesis previamente formulada".

La explicación científica tiene una importancia máxima en las ciencias, porque la ciencia está formada por teorías, y las teorías no son otra cosa que explicaciones racionales de determinados sectores de lo real.

Como escribe E. Nagel (1974, 17) "es el deseo de hallar explicaciones sistemáticas y controlables por elementos de juicio fácticos lo que da origen a la ciencia y es la organización y clasificación del conocimiento sobre la base de principios explicativos lo que constituye el objetivo distintivo de las ciencias".

172 EJERCICIOS

Dados los siguientes resultados obtenidos en el ejercicio 240 de la primera parte de esta obra:

Suicidios de varones en España entre los años 1965 a 1969: Media, 1236. Desviación típica, 44.8.

Suicidio de mujeres en España en el mismo tiempo: Media, 522. Desviación típica, 10.9.

Analizar e interpretar teóricamente los resultados y para ello:

A) Normalizar los resultados y establecer el índice de desigualdad entre ellos.

RESPUESTA

Para normalizar las desviaciones típicas es preciso hallar el coeficiente de variación según la fórmula siguiente:

$$\text{v. de hombres} = 44.8/1236 = .036;$$
$$\text{v.de mujeres} = 10.9/522 = .0208$$

Para medir la desigualdad entre ambas distribuciones se puede utilizar simplemente la razón o cociente entre las medias de suicidios de los hombres y las mujeres:

$$1236/522 = 2.36$$

B) Comparar entre sí los valores de cada distribución de frecuencias y una distribución con otra y formular los enunciados estadísticos.

RESPUESTA

En la comparación interna destacan la uniformidad de los valores que componen las dos distribuciones y, por tanto, podemos establecer que las variaciones anuales del número suicidio tanto en los hombres como en las mujeres son muy reducidas.

Dicha variación resulta cifrada cuantitativamente por el coeficiente de variación que es .036 en los hombres y .0208 en las mujeres. Lo que permite concluir que la variación anual sólo alcanza un 3.6% en los hombres y un 2.08% en las mujeres.

De la comparación entre ambas distribuciones se puede formular el enunciado estadístico de que el suicidio es más elevado en los hombres que en las mujeres en la proporción de 2.3 en hombres por 1 en mujeres.

C) Formular teóricamente las conclusiones estadísticas anteriores y efectuar su contraste con la teoría precedente.

RESPUESTA

De que el suicidio es mayor en los hombres que en las mujeres se puede inferir que el sexo influye en el sucidio.

Igualmente visto el suicidio, desde el punto de vista sociológico, como un índice de falta de integración social, se podría deducir que la integración social es menor en los hombres que en las mujeres.

Dada la escasa variación anual en el número de sucidios se podría caracterizar al suicidio como un fenómeno estable y regular en el tiempo.

El contraste de estas conclusiones con las teorías precedentes exigen el estudio de la sociología del suicidio y su comparación con los resultados de análisis similares.

173

Dada la siguiente tabla:

Respuestas	*% de respuestas autoritarias*	
	Clase media	*Clase trabajadora*
Piensan a favor de Mc Carthy	57.2	57.9
Piensan que siempre habrá guerras	59.7	58.7
Piensan que las mujeres deberían tener menos libertad	19.4	17.3

Respuestas a preguntas en tres encuestas (a hombres solamente) sobre autoritarismo, por clases sociales, en %, USA, año de 1965. Fuente: M. Rosenberg (1968, p. 87).

Analizar e interpretar teóricamente la tabla y para ello:

A) Explicar de qué tipo de tabla se trata, las variables que la forman y qué tipo de análisis es aplicable.

RESPUESTA

Es una tabla compuesta y resumida a la vez. Compuesta porque en realidad engloba tres tablas distintas, referentes a cada una de las respuestas por separado.

Resumida porque falta, en todos los caso, el % de respuestas no autoritarias, cuyo % es simplemente la diferencia entre el 100 % y el % dado de respuestas autoritarias.

Por ejemplo la tabla de la primera respuesta sería:

Respuesta	*Clase media*	*Clase trabajadora*
Piensa favorablemente a Mc Carthy	57.2	57.9
No piensa favorablemente a Mc Carthy	42.8	42.1
	100 %	100

B) Determinar la correlación entre las variables en cuestión y probar la hipótesis nula de que dicha correlación es significativamente distinta de cero.

RESPUESTA

La correlación, en el caso de todas las respuestas, la podemos hallar fácilmente mediante el coeficiente epsilón o diferencia de porcentajes entre la clase media y la clase obrera.

Epsilón es .7% , en la primera respuesta; 1% en la segunda y .1% en la tercera.

Hay, pues, en todos los casos una correlación insignificante en la muestra y, por tanto, se puede concluir que también con mayor razón en relación a la entera población de la que procede la muestra.

C) Formular los enunciados estadísticos y teóricos correspondientes.

RESPUESTA

El valor casi cero del coeficiente de correlación hallado autoriza a afirmar, en términos estadísticos, que no existe correlación entre las variables en cuestión. Esto implica que la relación entre ambas variables es, según la tabla considerada, simétrica.

Desde el punto de vista teórico esto significa que la clase social es independiente o no está relacionada con el autoritarismo medido por las respuestas de la tabla.

Sin embargo, esta conclusión contradice la derivada de otras investigaciones empíricas que han puesto de relieve una relación significativa entre ambas variables.

Por tanto, el análisis e interpretación de estos resultados se debe prolongar, introduciendo nuevas variables, en orden a descubrir si existe alguna variable, que encubran la relación entre dichas variables como se realiza en el ejercicio siguiente.

174

Dada la siguiente tabla en la que se prueba la relación entre las dos variables del ejercicio anterior, mediante la introducción de una nueva variable: la educación.

Respuestas en tres encuentas solo entre hombres a preguntas sobre autoritarismo según la clase social y la educación.

que da una respuesta autoritaria %:

	0-8 años de educación		*9-12 años de educación*		*12 o más años de educación*	
	Clase		*Clase*		*Clase*	
Respuestas	*Media*	*Obrera*	*Media*	*Obrera*	*Media*	*Obrera*
A favor de Mc Carthy	68.9	64.5	61.9	54.6	45.3	33.3
A favor guerra	71.4	59.7	68.8	62.2	48.2	38.4
En contra libertad mujeres	42.8	20	20	10.3	10.3	0

Analizar e interpretar teóricamente esta tabla y formular los enunciados estadísticos y teóricos correspondientes.

RESPUESTA

En esta tabla, que es resumida por los motivos señalados en el ejercicio anterior y además porque engloba tres tablas distintas una por cada nivel educativo, se relacionan las mismas variables que en el ejercicio anterior, con la diferencia de que al haber formado una tabla entre la clase social y el autoritarismo para cada nivel educativo distinguido, está controlada la variable educación.

La obtención de los coeficientes épsilons muestra que existe una cierta correlación entre las variables clase social y autoritarismo en todos los casos, que en algunos casos llega a un épsilon del 22.8

Admitida esta asociación, se trata de un tipo de relación asimétrica. En ella la variable clase social es una variable básica que lógicamente se ha de suponer que puede influir en el autoritarismo y no al revés.

Por tanto el análisis e interpretación de estas nuevas tablas obliga a llegar a una conclusión teórica opuesta a la del ejercicio precedente.

La variable clase social y el autoritarismo no son independientes. A la clase media, más elevada que la obrera, corresponde un mayor autoritarismo.

La explicación de esta contradicción se encuentra en la influencia de la variable educación en la relación entre las otras dos variables: la clase social y el autoritarismo.

Esta influencia actúa en el sentido de velar y encubrir en la tabla inicial la correlación entre la clase social y el autoritarismo.

Sin embargo, cuando se neutraliza esta influencia, controlando la variable edución, como se hace en la segunda tabla, entonces sale a la luz y se manifiesta la asociación existente efectivamente entre ambas variables, y se ve que en realidad este caso no desmiente lo hallado en otros estudios empíricos anteriores.

175

Dada la siguiente tabla:

Actitudes ante los derechos humanos y clase social.

Puntuaciones en aceptación de los derechos civiles	*Clase social*	
	Media	*Obrera*
Alta	27	45
Baja	73	55
	100 %	
	(120)	(120)

Fuente: Morris Rosenberg (1968, p. 94).

Analizar e interpretar teóricamente y para ello:

A) Señalar las variables y el tipo de la tabla y la clase de interpretación que procede.

RESPUESTA

Es una tabla de contingencia normal, en %, con dos variables dicotómicas, la clase social y la aceptación de los derechos humanos.

El tipo de interpretación aplicables es el referente a dos variables.

B) Determinar la correlación entre las variables de la tabla e interpretarla.

RESPUESTA

El coeficiente épsilon del −18% indica que existe una cierta correlación entre ambas variables. La relación entre ellas es, por tanto, asimétrica. En esta relación la variable básica es la clase social. Lógicamente hay que suponer que la correlación va en la dirección de la clase social a la aceptación de los derechos civiles.

Por tanto teóricamente esto significa que la clase social más baja presenta una mayor aceptación de los derechos civiles.

Sin embargo, ocurre como en el ejecicio anterior que, si se juzga críticamente esta conclusión, parece dudoso que a una clase social más elevada que generalmente supone una mayor educación intelectual y cívica pueda corresponder una menor aceptación de los derechos civiles.

175 (bis)

Dada la siguiente tabla en la que se estratifica la tabla anterior:

Aceptación de los derechos civiles según la clase social y la raza.

Aceptación derechos civiles	*Negros*		*Blancos*	
	C. media	*C. baja*	*C. media*	*C. baja*
Alta	70	50	30	20
Baja	30	50	70	80
	100 %			
	(20)	(100)	(100)	(20)

Fuente: M. Rosenberg (1968, p.105).

Analizar, interpretar y formular los enunciados estadísticos y teóricos correspondientes.

RESPUESTA

En la tabla se relacionan las mismas variables que en el ejercicio precedente, pero con la diferencia de que al comprender en realidad dos tablas distintas que relacionan la clase social y la aceptación de los derechos civiles, una para los negros y otra para los blancos, está controlada la variable raza.

La obtención de los coeficientes épsilons muestra ahora una correlación positiva entre clase social y aceptación de los derechos civiles del 20% en los negros y del 10 en los blancos.

La relación, pues, entre las dos variables en cuestión sigue siendo asimétrica y tampoco hay razón alguna para afirmar que la dirección de la asociación haya dejado de ser clase social → aceptación de los derechos humanos.

Sin embargo lo que ha cambiado es el signo de la asociación de negativo a positivo. Ahora a una clase social más elevada corresponde una mayor aceptación de los derechos civiles.

En consecuencia, la asociación entre las variables clase social y aceptación de los derechos civiles era engañosa en la primera tabla y producida por la variable incontrolada raza dado que esta tabla encubre que los negros son los que más aceptan los derechos civiles, como es natural dada la carencia que sufren de ellos, y los que pertenecen en mayor proporción a la clase obrera.

Por ello, cuando se controla o neutraliza la influencia de la variable raza, se manifiesta el carácter positivo de relación entre las variables en cuestión, supuesto lógicamente.

23. ANALISIS UNIVARIABLE

El análisis estadístico univariable o de las distribuciones de frecuencia presenta dos aspectos principales:

1. El resumen de las observaciones sobre una variable.

2. La comparación entre los valores de la misma o distintas variables, bien para expresarlas unas en función de otras o bien para medir la desigualdad que existe entre ellas.

Empezaremos por el primer aspecto.

RESUMEN DE LAS OBSERVACIONES SOBRE UNA VARIABLE

Las observaciones sobre una variable adoptan la forma de una distribución de frecuencias. Esta, como se ha indicado, representa la forma más simple de la tabulación. Proporcionan datos meramente descriptivos a diferencia de las tablas de dos o más variables que indican relaciones entre variables.

El análisis estadístico básico de estas distribuciones de frecuencias consiste en reducir el conjunto de observaciones sobre una variable a una medida típica que, como tal, se puede tomar como característica y representativa del grupo.

Esta reducción es útil y precisa, por razones operativas, expositivas, de síntesis y comparativas, y el valor que la expresa recibe el nombre de medida de tendencia central o promedio. Pero las medidas de tendencia central, como todos los resúmenes, son imperfectas y sólo se logran a base de una simplificación más o menos alejada de la realidad.

De aquí que como índice de la mayor o menor imperfección de estas medidas con relación a la distribución de frecuencias a que se refieren o, con otras palabras, de la dispersión real del conjunto de observaciones

que forman la distribución de frecuencias en cuestión respecto a esta medida de tendencia central, el análisis estadístico proporciona los llamados índices de variación.

Estos constituyen el complemento de las medidas de tendencia central o promedios. Reflejan el grado de dispersión de los valores de la variable que forman la distribución de frecuencias y señalan la mayor o menor aproximación con que la medida de tendencia central de que se trate, representa al grupo.

LOS PROMEDIOS Y LOS INDICES DE VARIACION

Las medidas de tendencia central y de dispersión más fundamentales empleadas en las distribuciones de frecuencias son las siguientes:

Distribuciones referentes a variables de escala nominal.

Promedio: Moda. Indice de variación: razón de variación, que es igual a

$$\frac{1 - \text{Frecuencia modal}}{\text{N o frecuencia total}}$$

Distribuciones referentes a variables de escala ordinal.

Promedio: Mediana. Indice de variación: Dispersión intercuartílica o recorrido interdecílico.

Distribuciones referentes a variables de escala de intervalo o de razón.

Promedio: Media aritmética o media geométrica.

Indice de variación: Desviación típica.

En relación con todos estos promedios e índices de variación, nos limitamos a presentar la anterior enumeración y prescindimos de exponer su teoría básica no solo por considerar que, normalmente deben ser conocidos por los que utilicen esta obra, sino también porque al tratar en el capítulo siguiente de las nuevas técnicas del análisis exploratorio tendremos ocasión de hacer referencia, sino a todos sí a algunos de ellos.

CRITERIOS PARA LA ELECCION DEL PROMEDIO

Para decidir qué promedio es más adecuado, en un caso concreto, moda, mediana o media, hay que tener en cuenta, además del tipo de escala de la variable ya indicado, el objetivo que se busca y el modelo de distribución de los datos.

Por ejemplo, en cuanto al objetivo que se busca, si se desea hallar el promedio del tamaño de las familias obreras españolas a efecto de un plan de construcción de viviendas sociales, entonces la moda sería un promedio más adecuado que la mediana y la media.

Similarmente, si, respecto al tipo de distribución, esta distribución no es simétrica, cuanto menos lo sea, la media pierde tipicidad. Así, en las curvas en U la media no existe prácticamente y por eso no se ajusta a la realidad.

Puesto que, como dice Mueller (1970, p. 142), muchos datos sociológicos –salarios, tamaños de ciudades, de familias, etc.– son fuertemente asimétricos, es esencial considerar el tipo de curva en la selección del promedio preferible.

LA DISPERSION RELATIVA Y EL COEFICIENTE DE VARIACION

La dispersión indicada por la desviación típica o por otro índice cualquiera de variación, se llama dispersión absoluta.

Sin embargo, una dispersión absoluta de 1.350 alumnos matriculados respecto a una matrícula anual de 17.000 como máximo, no tiene la misma significación e importancia que esa misma variación en relación con una matrícula anual de 8.000 alumnos.

De aquí la necesidad de conocer la importancia relativa de los índices de variación a fin de poderlos comparar con otros similares.

La medida de la importancia relativa viene dada por la dispersión relativa, cuya fórmula es la siguiente:

$$\text{dispersión relativa} = \frac{\text{dispersión absoluta}}{\text{promedio}}$$

En el caso de la media y la desviación típica, la dispersión relativa recibe el nombre de coeficiente de variación o de dispersión, cuya fórmula es la siguiente:

$$\text{Coeficiente de variacion} = V = \frac{\sigma}{\bar{x}}$$

σ es la desviación típica y $\bar{x}$ la media. El resultado se suele expresar en porcentajes.

También se puede medir la desviación de la media en unidades de desviación típica, z, normalizadas, independientes de las unidades empleadas. Da los resultados en todos los casos en unidades sigma, y son de gran utilidad en la comparación de distribuciones, aunque provengan de distintas poblaciones, siempre que éstas se distribuyen normalmente.

Su fórmula es:

$$z = \frac{x - \bar{x}}{\sigma}$$

En distribuciones normales, las puntuaciones z tienen una media de 0 y una desviación típica igual a 1.

Como derivaciones de las puntuaciones z, que tienen como finalidad principal eliminar los decimales y las puntuaciones negativas, (Cfr. Domenech, 1978, 580 y ss) se pueden citar:

a) las puntuaciones probit, de media igual a 5 y desviación típica igual a 1.

Su fórmula es: PROBIT = 5 + z.

b) Las notas T con media 50, desviación típica 10, y fórmula T = 50 + 10z, y

c) el coeficiente intelectual de Wechsler que tiene una media de 100 y una desviación típica de 15, según esta fórmula CI = 100 + 15z.

NORMALIZACION

En las distribuciones de frecuencias también es interesante, además de hallar sus promedios e índices de variación, encontrar una pauta para comparar fácilmente sus diversos valores.

Como dice Mueller (1970, p. 180 y ss.), una importante función del método estadístico es proporcionar las técnicas por las que las medidas individuales pueden ser interpretadas con propiedad o dos o más comparadas significativamente.

Existen muchos procedimientos al efecto, que pertenecen al grupo de operaciones de normalización, puesto que establecen una norma estadística apropiada en términos de la cual los datos brutos se pueden expresar y por ello son transformados en comparables.

Entre estos procedimientos destacan las proporciones, los porcentajes, las ratios, las tasas y los índices, además de las puntuaciones z acabadas de mencionar.

PROPORCIONES Y PORCENTAJES

La forma más simple de normalización son las proporciones en las que se comparan las frecuencias o valores de la distribución con N o el total de la misma.

Las proporciones, pues, expresan los valores de la distribución en términos o en función del valor general de la población, y no son otra cosa que las frecuencias relativas ya definidas.

Cuando se multiplican por cien, se tienen los porcentajes.

Las proporciones y porcentajes indican en cifras homogéneas la importancia relativa en el conjunto de cada frecuencia de los valores de la variable.

RATIOS O RAZONES, TASAS E INDICES

La razón de un valor A con relación a otro B, se define como A partido por B.

En términos aritméticos *ratio* es un vocablo genérico, que indica una comparación por división. Por ello cualquier persona puede construir tantas ratios como sus necesidades puedan sugerirle.

Tasas. Todas las tasas son "ratios". La tasa es la magnitud media de una variable por unidad de otra.

Básicamente una tasa es una combinación estadística de dos variables: la variable criterio, cuya variación se quiere saber, y la variable norma cuya variación deseamos eliminar. Esta última puede ser distinta.

Un segundo elemento en la construcción de la tasa es la elección de la base numérica estandar: 10, 100, 1000. Su función es eliminar los decimales para facilitar la cita y una más rápida comprensión.

La fórmula generalizada de una tasa es la siguiente:

$$\text{Tasa} = \frac{\text{Frecuencia de la variable criterio}}{\text{Frecuencia de la variable pauta}} \times \text{Base numérica}$$

$$\text{Tasa} = \frac{FC}{FP} \times BN$$

Las tasas pueden ser brutas o netas, según se calcule sobre la población total o sólo sobre la población que únicamente puede poseer la variable criterio, por ejemplo, en el caso de los nacimientos, las mujeres en edad de tener hijos. (Ver ejercicios 179, 180)

El índice. Por índice se designa:

1. La medida conjunta en la que se agrupan los distintos indicadores y elementos de una variable, ponderándolos según su importancia relativa.

2. La medida de los cambios de una variable a lo largo del tiempo, espacio, etc., en función del valor de la variable en un determinado punto del espacio, tiempo, etc, que se llama situación base.

3. La combinación de ambas nociones.

La situación base se suele valorar como 100.

Así un índice de coste de vida de 130 en 1970, sobre el año base de 1960, indica que el coste de la vida ha aumentado en un 30 por 100 sobre el año base considerado como 100. (Ver ejercicios 180 y 182)

ESTANDARIZACION

Así como un valor absoluto individual carece de significación a no ser referido a una apropiada pauta, igualmente una tasa o porcentaje individual posee escaso significado sociológico por sí solo. Lo adquiere sólo cuando se coloca codo a codo con tasas análogas a efectos de su comparación.

Por la estandarización se expresa la tasa de una categoría de la variable en función de los valores de la otra, es decir, suponiendo que su distribución con relación a la segunda variable es la misma que la de la otra categoría de referencia. (Ver ejercicio 181).

MEDIDA DE LA DESIGUALDAD

Las medidas de normalización sirven para comparar valores diversos de la misma o de distintas variables y, por tanto, sirven también para medir la igualdad o desigualdad que existe entre dichos valores, bien mediante su comparación o directamente.

La medida de la desigualdad es un punto no tratado frecuentemente en los manuales de Estadística social, a pesar del interés social y político que entraña.

Según Alker (1973, pág. 40) las medidas básicas de desigualdad individuales son las siguientes:

a) Las diferencias de valores y las ratios. Si se consideran dos individuos 1 y 2, la desigualdad entre ellos respecto a alguna variable se puede expresar bien restando los valores que detentan en dichas variables o por la razón o cociente entre dichos valores: $v_1 - v_2$ o bien $\frac{v_1}{v_2}$

b) La diferencia entre el valor de la variable que detenta un individuo y el valor medio para toda la población $v_1 - \bar{v}$

c) Por la relación de ventaja que consiste en la división de la diferencia entre un valor individual cualquiera v_i y el valor medio $\bar{v}$ dividido por el valor medio $\bar{v}$ o sea:

$$\frac{v_i - \bar{v}}{\bar{v}} = \frac{v_i}{\bar{v}} - 1$$

Esta ecuación da las diferencias entre valores y sus medias en proporciones de las medias respectivas y por ello hace posible la comparación de las desigualdades entre diferentes poblaciones.

A estas medidas se puede añadir el llamado por algunos (ver Informe Foessa, 1970, cuadro 20-5) "incremento relativo de prevalencia".

Su fórmula es $\frac{V_2 - V_1}{V_1 + V_2}$

Una aplicación de esta medida es el que A. de Miguel (Revista española de la Opinión Pública, n. 43, 1976) llama "indice de discriminación sexual", y que es igual a

$$\frac{\text{mujeres} - \text{varones}}{\text{mujeres} + \text{varones}}$$

También se puede consultar el artículo de Allison, P. D. (1978) *Measures of inequality*. "American Sociological Review".

LAS CURVAS DE LORENZ

Con relación a grupos de individuos se puede medir la desigualdad dividiendo el % de valores que detentan por el % de la población que representan.

En un eje de coordenadas se puede llevar el % de la población al eje de las X y el % de los valores al de las Y. Uniendo los puntos resultantes se forma la llamada curva de Lorenz, que muestra gráficamente el grado de desigualdad en la participación de valores.

El valor 1 de esta pendiente de la curva, que se produce en el punto en el que la ordenada y la abscisa son iguales, divide a la población en dos partes. Los que se encuentran en la curva a la derecha del punto reciben más de lo que debieran en condiciones de igualdad perfecta y los situados a la izquierda menos.

Por tanto se puede definir el coeficiente de desigualdad como la suma del % de población situado a la izquierda de dicho punto.

Este coeficiente mide el % de desfavorecidos en la población.

Para cuantificar el grado de desigualdad puesto de manifiesto por la curva de Lorenz se puede calcular el área que esta curva forma con la diagonal o línea de equidistribución.

Dividiendo esta área de desigualdad por su valor máximo se obtiene el coeficiente de Gini. Su fórmula es:

$$R = 1 - \sum_{i=1}^{n} (p_i - p_{i-1}) \cdot (q_1 + q_{i-1})$$

en la que p_i son las distintas proporciones acumuladas que representa la población de cada grupo social con relación al total de población de todos los grupos y q_i son las distintas proporciones acumuladas de participación por cada grupo social en el valor total disfrutado por toda la población, por ejemplo renta nacional, educación, etc.

El coeficiente de Gini mide aproximadamente la proporción que representa el área de desigualdad que forma la curva de Lorenz con la línea de equidistribución respecto al área de desigualdad máxima posible, que es la del triángulo formado por dicha línea de equidistribución y la ordenada y la abscisa principales a la derecha de la repetida línea.

La razón de concentración, que es el nombre que Gini da a su coeficiente (1953, p. 216) se puede obtener también por medio de un

planímetro, o sea midiendo planimétricamente el área de concentración efectiva y el área de concentración máxima y dividiendo la primera medición por la segunda. (Ver ejemplo 182)

EJERCICIOS

176

La media del coeficiente de inteligencia de una muestra de niños en edad escolar de alcance nacional ha resultado ser 90 y la desviación típica, 3. Hallar por medio de la tabla de las áreas de la curva normal el porcentaje de niños con un coeficiente de inteligencia entre 90 y 97 y entre 97 y 101.

RESPUESTA

Diferencia entre 97 y 90 = 7 : 3 (desviación típica) = 2,33.
Area de la curva normal para 2,33 = 0,4901 x 100 = 49,01 % .

Diferencia entre 101 y 90 = 11 : 3 = 3,66.

Area de la curva normal correspondiente a 3,6 (no hay valor para 3,66 en la tabla) = 0,4998409 – 0,4901 (área anterior) = 0,0097 = 0,0097 x 100 = 0,97%.

El coeficiente de los niños en cuestión está, pues, comprendido en un 49,01 entre 90 y 97, y en un 0,97% entre 97 y 101.

177

La media del peso de los soldados de una quinta es de 65 Kg. y la desviación típica de 4. Que% de soldados pesan de 75 a 80 Kg..

RESPUESTA

Se halla la diferencia entre 65 y 75 = 10, y 80 y 65 = 15.

Se dividen 10 y 15 por la desviación típica, 4, = 2,50 y 3,75.

Se busca en la tabla el área de la curva normal que corresponde a 2,5 y 3,75 sigmas. Es igual a 0,4938 y 0,4998.

La diferencia es el resultado buscado. Es = a 0,6 o un 6 % .

178

En una población la altura media de los mayores de 18 años es de1,60 m y la desviación típica 10 cm. Dentro de qué límites de altura se encontrará el 80 % de la población.

RESPUESTA

Dividiendo por dos se obtendrá un 40 % por encima y un 40 % por debajo de la media.

En la tabla el 40 % corresponde a un z de 1,29 aproximadamente.

Como la desviación típica vale 10 multiplicándola por 1,29, valor de un sigma, nos da 12,9 cm. por encima y por debajo de la media.

Luego 160 ± 12,9 = 147,1 y 179,9. Luego entre 147,1 y 172,9 está el 80 % buscado.

179

La difusión diaria media total de los periódicos en España en 1976 fué de 3.200.000 (supuesta). Cálcular la tasa por 1.000 de habitantes de difusión de los periódicos en España en 1977.

RESPUESTA

Fórmula:

$$\text{tasa} = \frac{\text{F C}}{\text{F N}} \text{ x } 1000$$

$$\text{tasa} = \frac{3.200.000}{37.000.000} \text{ x } 1000 = 86$$

Resultado: 86 periódicos al día por 1.000 habitantes.

180

Las retribuciones medias de las siguientes categorías socio-profesionales, fueron en 1976, las siguientes: (Ya, 15-4-77)

Peón	*319.000*
Especialista medio	*346.000*
Oficial medio (profesional)	*394.000*
Administrativo medio	*432.000*
Técnico no titulado medio	*496.000*
Titulado medio	*787.000*
Jefe de departamento	*1.081.000*
Jefe de Servicio	*1.228.000*

Tomando como base el sueldo del peón, forma un índice de escala de salarios en 1976.

RESPUESTA

Empleo	*Sueldo*	*Indice*
Peón	319.000	100
Especialista medio	346.000	108
Profesional oficio m.	394.000	123
Administrativo medio	432.000	135
Técnico no titulado medio	496.000	155
Titulado medio	787.000	246
Titulado sup. Jefe departamento	1.081.000	338
Titulado sup. Jefe de servicio	1.228.000	384

181

TASAS DE CRIMINALIDAD SEGUN GRUPOS DE EDAD Y RAZA

	Blancos			*Negros*		
Edad	*Población*	*Crímenes*	*Tasa*	*Población*	*Crímenes*	*Tasa*
15-24	27.000	378	14	5.000	65	13
25-44	36.000	324	9	4.000	32	8
45-74	27.000	108	4	1.000	3	3
	90.000	810	9	10.000	100	10

Estandarizar la tasa de crímenes de los negros sobre la población blanca. Interpretar los resultados.

RESPUESTA

Edad	*Población blanca*	*Tasa*	*Crímenes esperados en los negros*
15-24	27.000	13	351
25-44	36.000	8	288
45-74	27.000	3	81
	90.000		720

$$\text{Tasa estandarizada} = \frac{720}{90.000} \times 1000 = 8$$

Mientras que la tasa global de crímenes de los negros, según la primera tabla es de 10 por 1000 es decir, superior a la de los blancos: 9 por 1000, resulta que cuando se estandariza o uniforma esta tasa con la de los blancos, la tasa global de los negros es inferior en realidad a la de los blancos, para unos mismos niveles y distribución de población.

Por tanto, la tasa global de criminalidad más elevada de los negros, debe ser atribuída al mayor peso o proporción de los jóvenes en su población y no a una mayor propensión a la criminalidad (Basado en Mueller, J. H. (1970, pág. 190).

182

Según tabla contenida en la obra "Tres estudios para un sistema de indicadores sociales" de A. de Miguel y otros, publicada por la Fundación FOESSA, en 1967 la división en % de la población de España y la participación en % de dichos grupos en la enseñanza superior era la siguiente:

Categorías socioprofesionales	*% de la población*	*% de universitarios*
Colonos y obreros del campo	17	2,5
Capataces, obreros, peones y personal de servicios	38,5	4,8
Agricultores, propietarios, administración	22,8	5,1
Cuadros medios, empleados y vendedores	11,8	20,9
Empresarios y directores de empresas	7,2	26,9
Profesionales liberales, técnicos y afines	2,7	30,8
TOTAL	100,00	100,00

Se pide:

1. Formar la correspondiente curva de Lorenz.

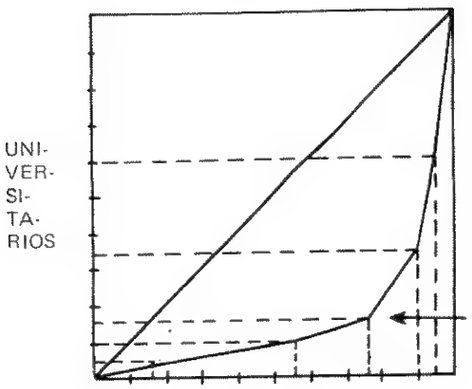
2. Hallar los coeficientes de desigualdad y el de concentración de Gini.

3. Hallar el índice en base 100 que exprese la participación proporcional de cada grupo en la educación universitaria.

4. Interpretar los resultados.

RESPUESTA

Curva de Lorenz



Frecuencias acumuladas

% acumulado Pob.	*% acum. Univ*
17,00	2,5
55,5	7,3
78,3	12,4
90,1	33,3
97,3	60,2
100,00	100,00

Punto de igualdad

Coeficiente de desigualdad = 78,3 % , suma de los porcentajes de los grupos de población que están en la curva a la izquierda del punto de igualdad.

Coeficiente de concentración de la educación universitaria de Gini. Aplicándo la fórmula se tiene:

p_i	q_i	$p_i - p_{i-1}$	$q_i + q_{i+1}$	$(p_i - p_{i-1}) \cdot (q_i + q_{i+1})$
0,17	0,025	–	–	–
0,555	0,073	0,385	0,098	0,04
0,783	0,124	0,228	0,197	0,04
0,901	0,33	0,118	0,457	0,05
0,973	0,602	0,072	0,935	0,06
1,000	1,000	0,027	1,602	0,04
		TOTAL		0,23

$$R = 1 - 0,24 = 0,77.$$

Indice de participación proporcional de cada grupo en la educación universitaria:

Grupos	*% de población*	*% de univ.*	*Indice*
Colonos y obreros campo	17	2,5	14,7
Obreros y peones etc.	38,5	4,8	12,4
Agricultores	22,8	5,1	22,3
Cuadros, empleados y vendedores . .	11,8	20,9	177,1
Empresarios y directores emp.	7,2	26,9	373,6
Profesionales liberales y técnicos . . .	2,7	39,8	1.474,0

El índice de concentración de Gini varía de 0 a 1. Por tanto, un valor de 0,74 indica un alto grado de concentración de la educación universitaria en unos sectores determinados de la población. Paralelamente, el coeficiente de desigualdad obtenido 78,3 % , señala que el % de los desfavorecidos en relación a la participación en la educación universitaria es muy elevado en relación al total de la población.

Igualmente el índice de participación proporcional señala que esta participación es muy desigual según los grupos de población, variando de un índice de 12,7 en los obreros a un índice de 1474 en las profesionales, es decir, en una proporción del 1 a 118.

24. ANALISIS EXPLORATORIO DE LAS DISTRIBUCIONES DE FRECUENCIAS DE TUCKEY

El análisis exploratorio es particularmente apropiado para el examen y estudio de las distribuciones de frecuencias, hasta tal punto que se puede decir que es en ellas donde tiene sus aplicaciones más importantes.

Por ello, dado que los procedimientos de análisis de estas distribuciones a que se acaba de hacer referencia y que se podrían llamar clásicos, deben haberse estudiado ya en Estadística, me limitaré aquí a exponer las aplicaciones básicas del análisis exploratorio de John W. Tuckey, para el estudio de las distribuciones de frecuencias, teniendo en cuenta la obra de B.H. Erickson y T.A. Nosanchuck, *Understanding data,* Mc. Graw-Hill, 1977.

Por otra parte, como señalan estos autores, las técnicas exploratorias se basan más no en lo que por lo general son deficientes los alumnos de las ciencias sociales: las matemáticas, sino en la curiosidad, la capacidad de observación, la imaginación, cualidades más propias y que atraen más a los estudiantes de ciencias humanas y sociales.

OBJETIVOS DEL ANALISIS EXPLORATORIO

El objetivo del análisis exploratorio es no sólo descriptivo, como en el caso de los promedios e índices de variación antes aludido, sino que también y sobre todo lo que pretende es estudiar las pautas y patrones que siguen los datos en las distribuciones, y las excepciones a estas pautas con el fin de encontrar pistas, sugerencias, ideas e hipótesis que tengan una significación teórica, sobre la realidad que describen numéricamente las distribuciones.

Se trata, pues, fundamentalmente, de hallar nuevas ideas, sugerencias e hipótesis, que puedan servir de base y orientación a nuevas investigaciones y al análisis de tipo confirmatorio, que busca verificar y contrastar hipótesis determinadas.

PROCEDIMIENTOS BASICOS DEL ANALISIS EXPLORATORIO

Particularidad de estas ideas, sugestiones e hipótesis es que se basan inmediatamente en la realidad, en los datos obtenidos de ésta, y que por tanto tienen una mayor base empírica que las hipótesis e ideas meramente especulativas e imaginativas.

El análisis exploratorio pretende, como señalan Erickson y Nosanchuk, descifrar profundamente los datos, no solo verlos sino también ver dentro de ellos.

Para ello utiliza dos procedimientos básicos: Primero, la diversa clasificación y ordenación de los datos, y segundo, la exploración de las características básicas de estas ordenaciones.

Adviértase que el análisis exploratorio no se basa en una única ordenación de los datos que constituyen la distribución, sino que cuando éstos lo permiten porque las unidades de observación que integran la distribución pertenecen a distintas categorías conocidas de una o más variables, se pueden y deben efectuar ordenaciones separadas de cada una de las categorías diferentes.

En principio cabe realizar con los datos en el análisis exploratorio todas las manipulaciones y ordenaciones que la imaginación sugiera.

Una vez realizadas las ordenaciones procedentes y de exploradas sus características, el análisis exploratorio actúa mediante el estudio y examen de los rasgos y particularidades que presentan las mismas, así como mediante las comparaciones de unas ordenaciones con otras.

Según los autores indicados, la regla general aquí es: elija el conjunto de números que atraiga más su interés. No pretenda verlo todo de buenas a primeras, libérese y concéntrese en lo que más llame su atención o más le interese.

LOS BATCHES

En este análisis se llama *batch* (lote, tanda, hornada) cada uno de los conjuntos de números referentes a una categoría de una o más variables.

Así las tasas de suicidios por naciones para hombres de 25 a 34 años se considera un batch. Cada una de estas tasas representa lo mismo, respecto a distintas naciones, que son en este caso las unidades de observación o de análisis.

Cada uno de los *batches* formados constituye el objeto inmediato del análisis exploratorio. Los *batches* del mismo tipo deben ser relacionados entre sí a efectos de realizar comparaciones entre ellos.

La formación de los *batches* requiere previamente la simplificación de los valores de la distribución de frecuencias. El procedimiento más simple al efecto es el redondeo, que puede efectuarse de varias maneras siendo la más común redondear al número entero inferior las décimas de .0 a .4; mientras que los decimales .5 a .9 se redondean al número entero próximo más alto.

Los elementos del *batch* son los tallos y las hojas *(stem and leaf).* El tallo está constituido por los dígitos principales correlativos que comprende el rango de una distribución y la hoja por el dígito siguiente. Por ejemplo en el valor 105, el tallo sería 10 y la hoja 5. El tallo y las hojas se disponen el primero a la izquierda y el segundo a la derecha, separados por una línea vertical.

CONSTRUCCION DEL BATCH

1. Se eligen los tallos. Para ello hay que ojear la distribución original y hallar su número más grande y más pequeño. Los tallos serán los dígitos relativos principales de este rango, considerándose como principales los que están a la derecha del dígito que se representa o se considera como unidad. Formado un tallo se puede cambiar por otro sin dificultad si se considera más adecuado, dado su carácter convencional.

2. Se ordenan los tallos. Es recomendable ordenarlos en sentido decreciente, de modo que los valores más elevados en cuantía sean también más altos en posición.

3. Se sitúa el dígito que corresponde a la hoja de cada valor de la distribución en el tallo que le corresponde.

4. Se escribe siempre al pie una nota indicando las unidades que representan el tallo y las hojas.

5. Finalmente se comprueba el *batch*. Para asegurarse de que no se ha omitido ningún valor se cuenta el número de hojas que debe igualar al número de valores de la distribución.

De este modo con toda facilidad y con un mínimo gasto de energía se obtiene una representación de los datos de la distribución que es gráfica, ordenada y completa.

CLASES DE BATCHES

En el caso de batches similares, referentes a las mismas unidades de observación pero a categorias de unas variables distintas, por ejemplo tasa de suicidio en hombre, por grupos de edad, podemos situar las hojas de un grupo de edad, p. ej. 25 a 34 a la derecha del tallo y las otras a la izquierda, 65 a 74. Este tipo de *batch* se llama *back-to-back,* espalda a espalda, y tiene gran valor práctico cuando se quiere comparar dos batches similares.

También se puede utilizar una forma compacta y otra detallada del *batch*. En la primera se colocan dos tallos en la misma línea separados por ejemplo con una coma. Para distinguir las hojas que corresponderían a cada tallo, se pueden separar con un signo, p. ej. dos puntos.

En la forma detallada, al reves, a cada tallo dedicamos dos líneas, y en el tallo superior situamos las hojas 5 a 9 y el inferior las hojas 0 y 4.

Finalmente, se pueden formar batches con más de una hoja;entonces las distintas hojas se separan con un signo, por ejemplo una coma.

Con estos batches la única información de las distribuciones originales que se pierde es la del nombre de las unidades de observación a que se refieren cada valor. Esto se puede remediar construyendo batches en los que las hojas en lugar del dígito final del valor, sean una abreviatura de la unidad de observación, u otro signo o número que las identifique.

CARACTERISTICAS DE LOS BATCHES

En los batches numéricos se pueden estudiar las cinco características principales siguientes:

1. Nivel, o magnitud del conjunto de número en su promedio.

2. Dispersión, o en qué medida los valores del batch se desvían del promedio.

3. Agrupación, o cómo se ordenan en grupos según los tallos del batch los valores de la distribución.

4. Forma, que adoptan los valores de la distribución, normal o no, simétrica o no simétrica.

5. Números inusuales, o sea valores mucho más elevados o más bajos que los demás.

La determinación de estas características nos da una imagen sintética y simplificada de cada conjunto, que facilita su comparación con otros.

EL NIVEL O PROMEDIO. Como nivel o promedio, aunque se puede utilizar la media, se recomienda emplear la mediana, porque es más resitente que la media, es decir porque resulta menos afectada por los casos extremos que pueden ser debidos al azar y en todo caso son mucho menos representativos del conjunto.

Sobre la mediana, si se tiene un número impar de casos hay un caso en medio de la distribución y éste es la mediana. Si el número es par, hay dos casos en medio y la mediana es el promedio de ambos.

La mediana es una cuartil. Para encontrar los otros dos se halla la mediana de las dos mitades alta y baja de la distribución.

Otro nivel o promedio fundado en la mediana es la TRIMEDIA o mediana media.

Su formula es $T_r i = (qs. + qi + 2Md) / 4.$

La trimedia utiliza más datos que la mediana al incluir el cuartil inferior y el superior, que proporcionan información adicional sobre la posición del centro evitando los extremos.

También se puede hallar la MEDIA CENTRAL.

Esta no es otra cosa que la media de los datos centrales del batch eliminados los valores que constituyen el cuartil inferior y superior.

El nivel o promedio proporciona la mejor predicción de la magnitud de los valores de la distribución.

DISPERSION. Las medidas de dispersión más adecuadas en el análisis exploratorio son:

El rango. Su fórmula es Xa – Xb, siendo Xa el valor más alto de la distribución y Xb el más bajo, es fácil pero no resistente.

La dispersión media o dq. Su fórmula es: dq = qs – qi, es decir, es la diferencia entre cuartil superior y el cuartor inferior. También se llama rango intercuartílico.

Media de las desviaciones absolutas. Es igual a la suma de las diferencias de los valores de la variable con la media, dividida por el número de casos N. La desviación típica es poco resistente, está afectada por los valores extremos, pero significa la misma cosa para todos los batches.

Tukey define también como medida de dispersión el paso, *step*, que es igual a 1.5 veces la dispersión media.

VALORES EXTREMOS. Con base en el paso, Tukey define como observaciones extrañas, *outsiders*, las que están a un paso por lo menos de la mediana por encima del cuartil superior o por debajo del inferior y muy extrañas, las que están a dos pasos por encima o por debajo.

Puesto que estos valores no son corrientes, Tukey sugiere que se noten los valores más alto y más bajo del *batch* que no sean o*utliers* y que el llama, adyacentes o cerca de ser *outliers.* Si no hay *outsiders* en el batch, los *adjacent* son sus valores extremos.

FORMA. La forma del *batch* resulta inmediata y directamente de la construcción del *batch.* La simple observación de la disposición que adoptan las hojas en el *batch* puede revelar si ésta es normal o no normal, simétrica o no.

En cada batch, Tukey consigna los valores extremos superior e inferior, los cuartiles también superior e inferior y la mediana, y N, o el número de datos del conjunto. (Ver ejercicios 183 y 186)

REPRESENTACION GRAFICA

Obedece a la necesidad de obtener grafos que retengan los datos importantes, sin detalles innecesarios, y puedan ser dibujados rápidamente.

Se trata de representar gráficamente los valores característicos y representativos del batch, acabados de exponer. El procedimiento consiste en situar estos valores síntesis en el papel tan alejados como lo estén en tamaño numérico. Para ello se necesita una escala que tenga el mismo tamaño que el tallo y las hojas.

La parte media central del *batch* es muy importante porque contiene gran cantidad de datos y los más seguros del conjunto. Para representarla se utiliza un estuche –*box*–, con líneas en los extremos del *box*, que marcan el cuartil superior y el inferior y una línea interior para señalar la posición de la mediana. Cuanto ésta no cae en medio del *box* indica asimetría.

Los valores extremos, *dots*, se representan el más alto y el más bajo por Xa y Xb, respectivamente y los que sin ser extremos son adyacentes más altos, por un circulito. (Ver ejercicios 183 y 186).

RESIDUALES

En el análisis exploratorio se puede operar no sólo con los datos originales sino con los llamados datos residuales, obtenidos mediante el cambio de variable o la standarización de los datos iniciales.

Se trata de proceder gradualmente en el análisis.

Primero se halla en ellos el promedio y la dispersión de los datos originales, y se analizan los datos originales.

Después se remueve por ejemplo la mediana, restando su valor de los datos iniciales, y se obtiene así un nuevo batch de residuales que puede ser objeto de nuevo análisis.

Luego se pueden remover el promedio y la dispersión de los datos iniciales standarizando éstos, obteniendo nuevos residuales, que también pueden ser objeto de análisis.

La forma inicial también se puede manipular mediante la transformación de los valores de la distribución.

Las transformaciones principales que se pueden efectuar con los valores x_i de la variable o distribución son las siguientes:

$$-1/x^2\ ;\ -1/x;\ \text{Log. x};\ x;\ x^2\ ;\ x^3\ ;\ \text{antilog x}$$

Las cuatro primeras corrigen la asimetría hacia la izquierda; y las cuatro últimas la asimetría hacia la derecha.

EJERCICIOS

183

Con el fin de comprobar empíricamente el fundamento matemático-estadístico de las muestras y dado el siguiente universo o población de seis elementos y sus medidas siguientes en la variable X

Población	*Medida en la variable X*
A	*2*
B	*3*
C	*4*
D	*6*
E	*9*
F	*12*

a) Formar todas las muestras distintas posibles de tres elementos que se puedan sacar de este universo y hallar la media de su valor en la variable X.

RESPUESTA

Para determinar el número de muestras posibles se puede utilizar la fórmula de las combinaciones que en este caso sería:

$$_6C_3 = \frac{6.5.4.}{3!} = \frac{6.5.4}{1.2.3} = \frac{120}{6} = 20.$$

Las muestras posibles, procediendo sucesivamente, serían las que se indican a continuación:

Muestras posibles n = 3.	Media (redondeada)
2, 3, 4	3
2, 3, 6	3.7
2, 3, 9	4.7
2, 3, 12	5.7
2, 4, 6	4.
2, 4, 9	5.
2, 4, 12	6.
2, 6, 9	5.7
2, 6, 12	6.7
2, 9, 12	7.7
3, 4, 6	4,3
3, 4, 9	5.3
3, 4, 12	6.3
3, 6, 9	6.
3, 6, 12	7.
3, 9, 12	8.
4, 6, 9	6.3
4, 6, 12	7.3
4, 9, 12	8.3
6, 9, 12	9.

b) Formar el batch correspondiente con todas las medias obtenidas de las muestras resultantes, tomando como tallo y hojas las que correspondan según dichas medidas.

RESPUESTA

Para formar el *batch* observamos que el rango de las medias obtenidas varía de 3 a 9 y que las cifras resultantes son unidades y décimas redondeadas. Por tanto, el tallo estará formado por las unidades y las hojas por las décimas, y el *batch* resultante será el siguiente:

10	
9	0
8	03
7	703
6	03703
5	3077
4	703
3	07
2	
1	
0	

184

Hallar, en su caso, y analizar las características del batch obtenido en el ejercicio anterior: nivel o promedio (utilizando la media), dispersión (utilizando la desviación típica), agrupación, forma y números o valores inusuales.

RESPUESTA

$$\text{Media} = \bar{x} = \frac{120}{20} = 6.$$

$$\text{Desviación típica} = \sqrt{\frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x}{n}\right)^2} = \sqrt{\frac{966}{20} - \left(\frac{120}{20}\right)} =$$

$$\sqrt{38.49 - 36} = \sqrt{2.49} = 1.57.$$

Agrupación: La inspección del batch nos muestra que la agrupación de la distribución es media, ni alta ni baja, en cuanto 9 de las 20 medias se concentran en los dos tallos más numerosos, el 5 y el 6.

Forma: Se ve que la forma es cercana a la normal. El tallo con hojas más numerosas es el de la media, el 6, y en los demás las hojas de los tallos disminuyen progresivamente según están más o menos alejados de la media. Se confirma, pues, la forma de distribución normal de todas las medias muestrales del mismo tamaño, obtenidas de la misma población, a pesar de tratarse en este caso de una población y una muestra reducidas casi al mínimo.

Números inusuales. Por la razón indicada de lo reducido de la población no hay números inusuales, es decir medias muestrales muy alejadas por exceso o defecto de la media de todas las medias o media de la población. Por ejemplo, las medias muestrales más alejadas, 3 y 9, no llegan a estar a dos desviaciones típicas o sigmas de la media de la población.

185

Dados los resultados de los ejercicios anteriores, analizarlos en relación a la teoría, del error muestral, comprobando, si conforme a esta teoría,

a) La media de las medias de la muestra $\bar{x}$, es igual a la media de la población $\bar{X}$.

b) La desviación típica de toda la población de las medias muestrales es igual al error muestral, $\sigma_{\bar{x}}$, según la fórmula de éste.

c) Si las medias muestrales se distribuyen conforme a la curva normal.

d) Qué % de las medias muestrales quedarán fuera de un nivel de confianza de 1 y 2 sigmas.

RESPUESTA

a) $\bar{X}$ población $= \frac{\Sigma x_i}{N} = \frac{36}{6} = 6.$

$\bar{X}$ de las $\bar{x}$ muestrales $\frac{\Sigma \bar{x}}{N} = \frac{120}{20} = 6$

Coinciden exactamente.

b) La desviación típica de las medias muestrales está hallado antes, ejercicio 247, y es igual a 1.57.

Error muestral. Hay que aplicar la fórmula para variables de intervalo y universos finitos siguiente:

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n} \cdot \frac{N - n}{N - 1}}$$

Por tanto necesitamos conocer σ^2 que es la varianza de la población inicial de la que se han obtenido las muestras.

Al tratarse de una población, la fórmula es

$$\sigma^2 = \frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x_i}{n}\right)^2$$

En cambio, si para determinar la desviación típica de la población que se designa por S^2 en lugar de σ^2, utilizamos una muestra de esa población, entonces se aplica la fórmula corregida siguiente:

$$S^2 = \frac{\Sigma x_i^2}{n-1} - \left(\frac{\Sigma x_i}{n-1}\right)^2$$

Utilizando, pues la primera fórmula se tiene:

$$\sigma^2 = \frac{290}{6} - \left(\frac{36}{6}\right)^2 = 48.3 - 36 = 12.3$$

Por tanto el error muestral, según la fórmula indicada sera:

$$\sigma_{\bar{x}} = \sqrt{\frac{12.3}{3} \cdot \frac{6-3}{6-1}} = \sqrt{4.1\,(.60)} = \sqrt{2.46} = 1.57.$$

Se ve también cómo el error muestral es igual a la desviación típica de las medias muestrales.

c) Ya indicamos al comentar en el ejercicio anterior la forma del *batch* que, según ésta, a pesar de la pequeñez de la población y de la muestra, la forma de distribución de las medias muestrales alrededor de la medida de la población se aproxima bastante a la de la curva normal.

d) El error muestral, en unidades sigma, vale en este caso 1.57. Por tanto el nivel de confianza de 1σ debe comprender todas las $\bar{x}$ muestrales comprendidas entre la $\bar{x}$ de la población $\pm\ 1\sigma$ ó 1.57, o sea $6 \pm 1.57 = 7.57$ y 4.43.

A este nivel de confianza, pues quedarán fuera las $\bar{x}$ de 7 muestras: 7.7, 8., 8.3, 9., 4.3, 3.7, y 3., es decir un 35% y dentro un 65% , este resultado es aproximado al % del área de la curva normal que abarca el nivel de confianza de 1σ, 68.7. En cambio, por la pequeñez de la población, al nivel de confianza de 2σ o sea superiores a 9.14 e inferiores a 2.8 no encontrando ninguna media muestral.

En conclusión, vemos prácticamente cómo la teoría matemática que sirve de fundamento a la muestra se confirma fundamentalmente aun en aún población y una muestra tan reducidas de 6. y 3.

186

Dada la siguiente tabla que expresa las tasas de natalidad por 1000 habitantes de las provincias españolas en el año de 1979 según en Anuario Estadístico de España, edición manual,

Total 16.06

Alava	*15.53*	*Albacete*	*17.07*	*Alicante*	*16.93*	*Almeria*	*19.76*
Avila	*11.14*	*Badajoz*	*16.62*	*Barcelona*	*14.00*	*Baleares*	*15.72*
Burgos	*15.56*	*Cáceres*	*13.62*	*Cádiz*	*21.04*	*Castellón*	*15.46*
Ceuta	*19.73*	*Ciudad R.*	*15.46*	*Córdoba*	*17.64*	*La Coruña*	*15.37*
Cuenca	*11.65*	*Gerona*	*17.05*	*Granada*	*17.52*	*Guadalajara*	*13.33*
Guipuzcoa	*12.62*	*Huelva*	*18.68*	*Huesca*	*11.22*	*Jaén*	*16.35*
León	*13.80*	*Lérida*	*14.16*	*Logroño*	*15.42*	*Lugo*	*12.17*
Madrid	*16.81*	*Málaga*	*18.51*	*Melilla*	*15.91*	*Murcia*	*20.66*
Navarra	*14.79*	*Orense*	*10.37*	*Oviedo*	*13.39*	*Palencia*	*14.14*
Las Palmas	*19.14*	*Ponteved.*	*17.02*	*Salamanca*	*14.70*	*Tenerife*	*16.53*
Santander	*16.44*	*Segovia*	*14.12*	*Sevilla*	*20.84*	*Soria*	*11.00*
Tarragona	*15.62*	*Teruel*	*11.10*	*Toledo*	*15.62*	*Valencia*	*17.21*
Valladolid	*16.61*	*Vizcaya*	*14.91*	*Zamora*	*11.92*	*Zaragoza*	*14.33*

SE PIDE, realizar el análisis exploratorio y para ello,

1º) Formar el batch correspondiente a esta tabla tomando como tallo las decenas y unidades y como hojas las décimas redondeadas.

2º) Determinar sus características: nivel o promedio, dispersión, agrupación, forma y números inusuales y su representación gráfica.

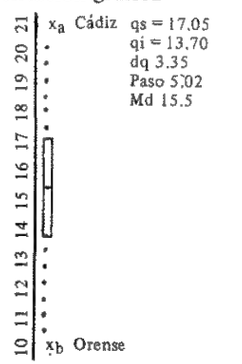
RESPUESTA

1º) Para formar el *batch*, es conveniente eliminar primero las centésimas, redondeándolas en las décimas. Hechos esto, el rango de la distribución o tabla varía 10.4 a 21., luego se pueden usar como tallos las decenas y unidades y como hojas las décimas.

El *batch* numérico y nominal resultante serían los siguientes:

22		22	
21	0	21	CA.
20	78	20	MU.SE.
19	817	19	AL.GC.CE.
18	75	18	H.ML.
17	161502	16	AB.CO.GE.GR.PO.V.
16	9648546	16	A.BA.J.M.TF.S.VA.
15	575544669	15	VI.PM.CS.CR.C.LO.T.TO.ML.
14	02817193	14	B.L.NA.P.SA.SG.BI.Z.
13	66384	13	BU.CC.GU.LE.O.
12	62	12	SS.LU.
11	172019	11	AV.CU.HU.SO.TE.ZA.
10	4	10	OR.
9		9	

Representación gráfica



Promedio: Se utiliza la mediana como es corriente en el análisis exploratorio. En las distribuciones de frecuencias, como ésta, sin agrupar, la mediana es la categoria que las divide en dos partes iguales. Si el número de casos es impar, hay, cuando se ordenan según su frecuencia de mayor a menor, un caso medio y este es la mediana. Si es par, hay dos y la mediana es la mitad de ambos. En el ejercicio tenemos 52 casos o provincias. Luego habrá dos provincias en medio de la distribución, que ocupan sus lugares 26 y 27, y tienen como valor las dos 15.5 y 15.5, luego la mediana o mitad de ambas será también 15.5.

Dispersión: La que corresponde es la desviación cuartílica dq. Para determinarla, se hallan de la forma indicada las medianas de cada una de las dos mitades de la distribución es decir de las 26 provincias con tasas más elevadas y las 26 con tasas más bajas.

La diferencia entre estas dos medianas o cuartiles es la desviación cuartílica, dq.

En este caso sería igual = dq = 17.5 – 14.70 = 3.35

Agrupación: Se puede considerar relativamente alta en cuanto de los 12 tallos del *batch*, en cinco seguidos del mismo, se concentra el 84 % de los casos de la distribución.

Forma: Representa una cierta similitud con la forma de la curva normal, si bien destacan algunas irregularidades en relación a la misma.

Números inusuales: No los hay. Los valores extremos los tienen Cádiz y Orense.

187

Interpretar los resultados obtenidos en el anterior ejercicio y exponer las sugerencias que se crean se deriven del análisis efectuado.

RESPUESTA

Aunque la disparidad de las distintas provincias españolas en cuanto a sus tasas brutas de natalidad no sea extrema, sin embargo se pueden distinguir en ellas tres grupos diferenciados.

1) Un pequeño grupo de provincias con tasas brutas de natalidad relativamente altas, formado por provincias del sur de España: Cádiz, Sevilla, Málaga, Murcia, Gran Canaria, Almeria, Huelva y Ceuta.

Este resultado es congruente con la conclusión de Jordi Nadal (La población española. Ariel, 1970, p. 243) según la cual las tasas provinciales de fecundidad matrimonial y tasas de reproducción "muestran que hay una España más vital al sur de Toledo, con sus ápices en Cádiz, Andalucia Penibética y Murcia, y una España menos vital de Toledo para el Norte, más diversificada que la primera".

2) Un segundo pequeño grupo con tasas relativamente inferiores a las de las demás provincias, y que se puede llamar deprimido demográficamente. Forman parte de él las provincias de Orense, Avila, Cuenca, Soria, Teruel y Zamora.

3) El tercer grupo incluye todas las demás provincias españolas, es decir, la mayoría de las mismas, con oscilaciones y diferencias entre ellas relativamente moderadas y en relación a las tasas brutas para toda la nación.

Como sugerencias sobre nuevas investigaciones sería preciso, en primer lugar, un estudio detenido y específico de los factores que pueden explicar la mayor vitalidad de las provincias del sur.

En segundo lugar debería también realizarse una investigación similar sobre los factores que son causa, en las provincias deprimidas demográficamente de su situación.

25. ANALISIS BIVARIABLE

CORRELACION DE VARIABLES

Según su etimología, correlación designa una relación recíproca. Sin embargo, en Estadística señala no una relación recíproca en sentido estricto, sino la correspondencia en las pautas de variación de dos o a veces más de dos variables.

Si dos variables están relacionadas de tal modo que cuando aumentan o disminuyen los valores o frecuencias de una variable, aumentan o disminuyen los de la otra, se dicen que están correlacionadas positivamente. Al contrario, si al aumento en una variable corresponde una disminución en la otra, la correlación es negativa.

El segundo tipo de aplicación básica de la estadística, en el análisis de los resultados de la investigación social, es la determinación numérica del grado y signo de la asociación entre dos variables.

Aquí no se trata ya, como en los promedios e índices de variación, de distribuciones de frecuencias, o tablas referentes a una sola variable, sino de cuadros que relacionan dos variables.

Estos cuadros pueden adoptar dos formas: la llamada distribución bivariante conjunta y la que se conoce con el nombre de tabla de contingencias.

La distribución bivariante conjunta consiste en una tabla que expone, respecto a cada individuo objeto de observación, los valores de dos variables distintas referentes a cada individuo.

La tabla de contingencias consiste en cruzar en un cuadro o tabla las varias categorías, dimensiones o valores de dos variables, consignando en cada casilla de la tabla el número de observaciones que reunen a la vez las dos dimensiones, categoría o valores de las variables que se cruzan o corresponden en la casilla de que se trata.

DISTRIBUCION BIVARIANTE

Individuos	*Peso*	*Talla*	*Sexo*
1	78	1,60	
2	65	1,58	Hombre
3	80	1,84	Mujer

TABLA DE CONTINGENCIAS

	Estado civil		
	Casado	*Soltero*	*Viudo*
Hombre	58	65	72
Mujer	42	85	94

COEFICIENTES DE CORRELACION

Los coeficientes de correlación o de asociación son índices numéricos, obtenidos mediante determinadas operaciones estadísticas, que cifran cuantitativamente el grado y, en su caso, el signo de la correlación entre dos variables.

Según su naturaleza, constituyen razones en las que se comparan por división una medida de la correspondencia o divergencia en la variación de las variables (suele ser la reducción proporcional del error de predicción de una variable, debida al hecho de tener en cuenta otra nueva variable) con el valor máximo que puede alcanzar dicha asociación o divergencia. Varían, por tanto de 0 a 1 o de -1 a + 1, cuando pueden tomar valores negativos.

De la simple inspección y lectura de una tabla, si es en porcentajes y no muy extensa y complicada, se puede deducir si existe una asociación en la variación de dos variables y el sentido de ésta. La función de la estadística en este punto, como en el anterior, es proporcionar a la sociología técnicas para cifrar matemáticamente el grado de asociación entre estas variables. Estas técnicas constituyen los llamados coeficientes de asociación o correlación.

Ya vimos cómo uno de los tipos de hipótesis son las hipótesis de dos variables que postulan una relación de asociación o variación conjunta entre dichas variables.

Con relación a estas hipótesis, que son sin duda las más frecuentes en la investigación social, los coeficientes de asociación nos dicen cuantitativamente el grado de asociación de las variables en cuestión y por tanto de ellos se puede inferir en qué grado la investigación realizada corrobora las hipótesis formuladas.

Además los coeficientes de asociación son instrumentos de predicción. Si observamos que dos variables están asociadas en su variación, y si vemos después que una de ellas aumenta o disminuye en una comuni-

dad, podremos prever, dada la asociación en la variación descubierta, que la otra también aumentará o disminuirá.

Por ejemplo, suele haber en las investigaciones una fuerte asociación entre las variables renta y consumo, cosa que por lo demás es de experiencia general. Por tanto, siempre que aumente la renta en una comunidad, podemos prever o esperar un aumento en el consumo.

Los coeficientes de asociación por sí solos no indican influencia o causalidad, es decir que una variable influya o sea causa de la otra, pero constituyen una de las condiciones para que se pueda establecer una relación de causalidad entre las variables en cuestión.

Otras condiciones son a) la prioridad en el tiempo de la variable causa y b) la no existencia de otras variables que influyan a su vez en la asociación detectada entre las variables en cuestión.

Sin embargo, en la práctica, es muy raro que se pueda estar totalmente seguro de esto último, por lo que se admite que se puede estimar que existe una relación de causalidad entre dos variables si están asociadas y si se dispone de una explicación teórica, dentro de la cual la influencia de una variable en otra aparece como racional y plausible.

En cuanto a la significación de los coeficientes de asociación, aunque no existe una norma válida para todos los casos, pues tal significación, no depende sólo de su tamaño, se admite que su significación, en principio, es la siguiente:

	+	.70	muy fuerte
+ .50	a	.69	sustancial
+ .30	a	.49	moderada
+ .10	a	.29	baja
+ .01	a	.09	despreciable

CLASIFICACION DE LOS COEFICIENTES DE CORRELACION

Se pueden clasificar, como se hace en el cuadro que sigue, según el número de variables, su dirección, el tipo de variables a que se aplican y sus requisitos.

TIPOS DE COEFICIENTES SEGUN

- El número de variables
  - Simples, entre sólo dos variables
  - Múltiples, entre más de dos variables
  - Parciales, entre dos o más variables, controlando o manteniendo constante otra u otras variables
- Dirección
  - Simétricos. Su valor no cambia si se cambia la variable considerada independiente
  - Asimétricos. Cambian de valor con el cambio de variable independiente
- Tipo de variables a que se aplican
  - Dicotómicas las dos: Q. de Yule; Phi
  - Las dos nominales: Lambda; Tau-y
  - Las dos ordinales: Tau; d, de Sommers; Gamma; Rho Spearman
  - Una V. dicotómica y otra de intervalo o razón: Biserial; Biserial puntual; Tetracórica
  - Las dos de intervalo o de razón: R de Pearson; Rho si los casos son menos de 30
  - Una variable nominal y otra ordinal: Zeta de Wilcoson
  - Una variable nominal y otra de intervalo: Etha
  - Una variable ordinal y otra de intervalo: Multiserial de Jaspers
  - Una variable dicotómica y otra ordinal: Biserial para rangos
- Requisitos
  - Paramétrcos: Exigen variables de intervalo y distribución normal de la población
  - No paramétricos: No exigen estos requisitos

EL COEFICIENTE EPSILON

El coeficiente epsilon es un coeficiente de asociación muy simple y poco riguroso, pero es muy fácil de calcular, por lo que resulta muy útil y práctico en los casos que sea necesario obtener una mayor exactitud.

Este coeficiente sólo es aplicable a porcentajes. Por tanto, en el caso de tablas con números absolutos no se puede aplicar, salvo que primero se expresen los valores de dicha tabla en porcentajes.

Consiste simplemente en la diferencia de tantos por cientos entre las columnas de la tabla, dos a dos. Hay que tener muy en cuenta que esta diferencia es necesario hallarla en sentido contrario al que hayan sido hallados los porcentajes.

Por ejemplo, si los porcentajes han sido hallados en sentido vertical, como es normal según la regla de Ziesel, el coeficiente epsilon se halla en sentido horizontal, restando horizontalmente, dos a dos los porcentajes de las columnas. (Ver ejercicio 188)

EL COEFICIENTE Q DE YULE, Y EL COEFICIENTE PHI

Ambos coeficientes son apropiados para medir la asociación de variables de escalas ordinal o nominal.

Unicamente son aplicables en tablas de 2 x 2, es decir, de dos columnas por dos filas.

La fórmula del coeficiente Q de Yule es la siguiente:

$$Q = \frac{a \cdot d - b \cdot c}{a \cdot d + b \cdot d}$$

a	b
c	d

La notación del coeficiente phi es la misma, pero la fórmula no.

Es la siguiente:

$$\varphi = \frac{a \cdot d - b \cdot c}{\sqrt{(a + d)\ (c + d)\ (a + c)\ (b + d)}}$$

Hay que tener en cuenta que la Q de Yule no es aplicable cuando una de las casillas está vacía, o tiene como valor cero, pues da entonces falsamente un resultado de más o menos 1.

Con relación a los signos dos coeficientes, es de advertir que si las dos dimensiones de las variables indican una idea de orden, como clase alta – clase baja, deben conservarse los signos; si no la hay, no es necesario. En el primer caso, las dos variables deben ordenarse de manera homogénea.

El coeficiente φ no siempre varía de + 1 a – 1.

En concreto, el valor máximo que puede alcanzar este coeficiente viene dado por la siguiente fórmula:

$$\varphi \text{ máximo} = \sqrt{\left(\frac{p}{q}\right) \cdot \left(\frac{p'}{q'}\right)}$$

a	b	p
c	d	q
p'	q'	

(Ver ejercicios 191 y 193)

COEFICIENTE LAMBDA

Los coeficientes de Yule y phi sirven para medir la asociación entre variables nominales, pero únicamente cuando ambas son dicotómicas, es decir, tienen solo dos categorías.

Otro coeficiente apropiado para variables de escala nominal que no tiene la limitación de Q y phi es el coeficiente lambda de Goodmann y Kruskal.

Según Freeman (1971, p. 102 y ss), la asociación se puede considerar como un problema de predicción. Así, dos variables están asociadas si el conocimiento de una de ellas sirve para mejorar la predicción de los valores de la otra.

El coeficiente lambda se funda precisamente en este carácter predictivo de los coeficientes de asociación.

Veámoslo con un ejemplo de Freeman. Supongamos que se propone un problema de lógica a 50 estudiantes universitarios y obtenemos los siguientes resultados:

Resolvieron el problema	30
Fallaron	20
	50

Dado este resultado, la resolución del problema representa la mejor predicción del mismo. Es la moda y ésta como es sabido representa la mejor predicción en el caso de variables nominales.

Esto puede comprobarse si valoramos las dos posibles predicciones. Si predecimos que cada individuo ha resuelto el problema, 20 de las 50 predicciones serán falsas.

Si por el contrario, predecimos que todos fallaron, 30 predicciones de las 50 serán falsas. Escogiendo, pues, la clase modal, cometemos menos errores.

Supongamos de nuevo que se nos facilita información sobre cada estudiante referente a la variable de haber seguido o no un curso de matemáticas.

Esta variable se puede llamar independiente, pues nos puede ayudar a predecir los valores de la primera.

Sea esta tabla:

	Siguieron curso de matemáticas	*No lo siguieron*	*Total*
Resolvieron el problema	22	8	30
No lo resolvieron	3	17	20
	25	25	50

Con esta información podemos valorar nuevamente el acierto en la predicción de la resolución o no del problema de lógica, teniendo en cuenta el estudio o no de matemáticas.

Si ambas variables están asociadas, al hacerlo así reduciremos los errores en la predicción cometidos inicialmente.

Primero consideremos los estudiantes con preparación matemática. De ellos 22 resolvieron el problema y tres no. La predicción elegida debe ser, pues, la de que resolvieron el problema con lo que cometemos un error de 3 en 25.

Por el contrario, para los estudiantes sin preparación de matemáticas, la moda sería fallaron y por tanto la mejor predicción. El error que cometemos con ella es el de 8 en 25.

Volvamos al principio. Como vimos el error que cometíamos en la predicción sin tener en cuenta el estudio de matemáticas será de 20 en 50.

Pero ahora cometemos solo tres errores en el grupo con formación matemática y 8 en el de sin preparación matemática. En total 11 errores en los dos grupos.

Por tanto, tener en cuenta la formación matemática nos ha permitido reducir los errores de predicción de 20 a 11.

Esta reducción representa pues:

$$\frac{20 - 11}{20} = \frac{9}{20} = 0{,}45 \text{ ó } 45\ \%$$

Este 45 % es el coeficiente lambda que nos indica el % en que la variable independiente nos sirve para mejorar nuestra predicción, y que, mide el grado de asociación entre ambas variables.

La fórmula para hallar lambda cualquiera que sea el número de dimensiones de las variables es la siguiente:

$$\lambda_a = \frac{\Sigma fi - Fd}{N - Fd}$$

En donde Σfi = a la suma de frecuencias máximas de las correspondientes a la variable independiente en cada columna.

Fd = a frecuencia máxima correspondiente a los totales de la variable dependiente.

N = al número de casos de la tabla, o su total general.

Hay que tener en cuenta que estadísticamente es posible calcular lambda nuevamente tomando como variable independiente a la dependiente considerada primero. Entonces claro es que hay que hallar las frecuencias máximas en dirección contraria.

Igualmente se puede hallar lambda, haciendo caso omiso de cual es la variable independiente y cuál la dependiente.

Entonces la fórmula es:

$$\lambda_b = \frac{\Sigma fr + \Sigma fc - (Fr + Fc)}{2N - (Fr + Fc)}$$

En la que fr es la suma de frecuencias máximas dentro de las filas.

fc = a la suma de las frecuencias máximas dentro de las columnas.

Fr = a la frecuencia máxima entre los totales de las filas.

Fc = a la frecuencia máxima entre los totales de las columnas.

Aplicando estas fórmulas en el ejemplo utilizado en la explicación se tiene:

$$\lambda_a = \frac{22 + 17 - 30}{50 - 30} = \frac{9}{20} = 0{,}45$$

en el caso de la formación en matemáticas como variable independiente.

Variable independiente resolución del problema:

$$\lambda_b = \frac{(22 + 17) - 25}{50 - 25} = \frac{14}{25} = 0{,}56$$

Sin considerar variable independiente:

$$\lambda_{ab} = \frac{(22 + 17) + (22 + 17) - (25 + 30)}{2\,(50) - (25 + 30)} = \frac{23}{45} = 0{,}51$$

Según Goodman y Kruskal (1979, 743) lambda es 1 siempre que la población se concentra en las casillas de modo que nunca dos de las casillas modales se encuentran en la misma fila o columna, sino en la diagonal de la tabla. Sin embargo, a veces el coeficiente 1 derivado de esta posición puede indicar una asociación exagerada.

Igualmente, cuando las casillas modales se encuentran en la misma fila o columna el coeficiente será cero, aunque a veces puede ocurrir, en este caso, que exista una cierta correlación entre las variables no detectada por este coeficiente.

Este coeficiente depende de los marginales o totales de la tabla. Con esto se quiere decir que no solo depende de las frecuencias de las casillas interiores de la tabla, sino también de la proporción que guarden entre sí los totales o marginales de cada categoría de las variables de la tabla. (Ver ejercicios 189 y 190)

EL COEFICIENTE DE ASOCIACION GAMMA DE GOODMAN Y KRUSKAL

Este coeficiente de correlación se emplea para medir la asociación entre variables con escala ordinal, es decir, aquellas en las cuales las dimensiones o categorías de las variables expresan un orden entre ellas, pero no intervalos iguales.

El fundamento de gamma es similar al acabado de indicar de lambda.

Para hallarlo se computan las concordancias y las inversiones en las ordenaciones que representan las dos variables y se aplica luego la siguiente fórmula:

$$\lambda = \frac{\text{número de concordancias} - \text{número de inversiones}}{\text{número de concordancias} + \text{número de inversiones}}$$

El número o frecuencia de concordancias se puede representar por fc y el de inversiones por fi.

Gamma es indeterminado si la población se halla enteramente en una sola fila o columna de la tabla. Es uno si la población se concentra en la diagonal principal y -1 si se concentra en la diagonal secundaria. Gamma es cero, en el caso de independencia o no asociación, pero lo inverso no siempre es verdad.

Si gamma varía entre 0 y 1 indica la probabilidad de acierto de la predicción del orden de una variable partiendo del orden en la otra. Si varía de 0 a -1 señala un orden inverso entre las dos variables.

Para calcular el número de concordancias e inversiones hay que distinguir dos casos. El primero el de las distribuciones bivariantes conjuntas y el segundo el de las tablas de contingencias.

Supongamos en relación al primer paso que se ha pedido a dos jueces que ordenen a cinco personas según su civismo. Sea esta la ordenación resultante:

Persona	*Juez H*	*Juez I*
A	4	4
B	1	2
C	3	3
D	2	1

Esta ordenación se puede presentar según el orden de mayor a menor puntuación asignada por el juez H. Entonces tendríamos el siguiente orden:

Persona	*Juez H*	*Juez I*	*Nº inversiones*	*Nº de concordancias*
A	4	4	0	0
C	3	3	0	0
D	2	1	0	2
B	1	2	1	2

Para hallar las inversiones y concordancias se parte del pie de la escala. Se cuentan como inversiones en la escala del juez I respecto a la escala del juez H todas las puntuaciones inferiores a la considerada que se encuentran encima de ella.

Como concordancias se cuentan todas las puntuaciones superiores que existan por encima de la considerada.

En el ejemplo propuesto se tendría:

$$\gamma = \frac{5-1}{5+1} = \frac{4}{6} = 0{,}66$$

En el caso de tablas de contingencias, la frecuencia de las concordancias, fc, es igual a la suma de los productos de los valores de cada casilla por la suma de los valores de las casillas que cada casilla tiene a la vez debajo y a la derecha.

La frecuencia de las inversiones, fi, es igual a la suma de los productos de los valores de cada casilla por la suma de los valores de las casillas que cada casilla tiene a la vez debajo y a la izquierda. (Ver ejercicios 189 y 194)

COEFICIENTE RHO DE SPEARMAN

Otro procedimiento para hallar la asociación entre variables ordinales es el coeficiente Rho de Spearmann.

También se puede utilizar para variables de intervalo con un número de casos relativamente pequeño, menos de 30, en lugar del coeficiente r de Pearson. Es más fácil de calcular que éste, pero menos satisfactorio matemáticamente.

Supongamos que los siguientes individuos han obtenido las siguientes puntuaciones en autoritarismo y prestigio social:

Individuos	*Autoritarismo*	*Prestigio social*
A	50	9
B	30	1
C	60	5
D	20	6
E	80	4

Estas puntuaciones se pueden ordenar por rangos. La más elevada será el rango 1 y así sucesivamente. Entonces tendríamos:

Individuos	*Rango en Ant.*	*Rango en Pr.*	*D*	D^2
A	3	1	2	4
B	4	5	1	1
C	2	3	1	1
D	5	2	3	9
E	1	4	3	9
				24

Determinados los rangos, el cuadrado de sus diferencias nos da una medida de las divergencias entre ellos.

Para hallar su significación se compara con el valor máximo posible de las divergencias en el orden de los rangos según la siguiente fórmula:

$$\rho = 1 - \frac{6 \cdot \Sigma\, d^2}{N \cdot (N^2 - 1)}$$

Aplicando la fórmula al ejemplo se tiene:

$$\rho = 1 - \frac{6 \cdot 24}{5 \cdot (25-1)} = \frac{144}{120} = 1 - 1{,}20 = -0{,}20$$

Existe una asociación baja negativa entre ambas variables.

Este coeficiente plantea el problema del caso en que dos o más unidades de observación reciben la misma puntuación respecto a la misma variable.

Si ocurre así, a cada puntuación igual se le asigna el rango promedio de los que se hubieran obtenido para estos valores si no hubieran existido empates. (Ver ejercicios 189 y 195)

EL COEFICIENTE DE CORRELACION r DE PEARSON

Este coeficiente es aplicable al caso de variables con escalas de tipo intervalo o de razón, es decir referentes a variables cuyos valores son distintos, representan un orden y tienen intervalos iguales, y, en el caso de escalas de razón con origen natural cero.

Se funda, como el coeficiente lambda, en la proporción en que reduce el error de predicción la introducción de una nueva variable en relación con el error original de la primera.

Adviértase que r^2 recibe el nombre de coeficiente de determinación: r^2 es igual a la proporción de la varianza de y que puede predecirse a partir de x y $1 - r^2$ es la proporción de la varianza que no puede predecirse en la variable. Reciben el nombre de varianza explicada y varianza no explicada.

r^2 sólo puede ser positivo. En cambio su raiz cuadrada, r, el coeficiente producto-momento de Pearson, se le pueden atribuir signos y hay fórmulas que dan el signo directamente, como la siguiente:

$$r = \frac{\Sigma xy - \frac{(\Sigma x) \cdot (\Sigma y)}{N}}{\sqrt{\Sigma x^2 - \frac{(\Sigma x)^2}{N} \cdot \Sigma y^2 - \frac{(\Sigma y)^2}{N}}}$$

Hay que tener en cuenta que aunque r^2 y r son interdependientes cada uno tiene su significación peculiar.

r^2 mide la proporción de la variación total de una variable que está asociada o es explicada por otra. En otras palabras mide la influencia ejercida por una variable sobre la otra.

Por el contrario, r tiene un carácter dinámico y mide la tasa de cambio de una variable con relación a otra. Por ello r es un instrumento predictivo. (Ver ejercicio 196).

EL COEFICIENTE ETHA

El empleo del coeficiente de correlación exige que se cumplan ciertos requisitos en las distribuciones de las variables a correlacionar.

El primero de ellos, ya indicado, que estas distribuciones sean de tipo de intervalo o de razón.

El segundo que la curva que formen las distribuciones sea bastante simétrica y unimodal.

El tercero que la asociación en la variación sea por lo menos aproximadamente rectilínea, o sea que la ecuación de regresión de ambas variables pueda representarse por una línea recta.

Hay muchos casos en que esta última condición no se cumple porque no es lineal sino curvilineal la asociación en la variación de las variables. En estos casos no se debe aplicar r y en su lugar hay que utilizar otro coeficiente, el etha.

El fundamento de este coeficiente es similar al de Pearson. Combina como éste las varianzas de las variables correlacionadas para obtener el porcentaje de reducción en el error de una variable producido por el hecho de tener en cuenta la segunda variable.

La fórmula de este coeficiente es la siguiente:

$$\eta = \frac{\sum_{j=1}^{k} n_j \, (\bar{Y}_j - \bar{Y})^2}{\Sigma \, (Y_i - \bar{Y})^2}$$

n_j = número de observaciones del subgrupo j de x.
$\bar{Y}_j$ = media del subgrupo j.

$\overline{Y}$ = media aritmética global.

k = número de subgrupos.

Y_i = valor de la variable de intervalo.

N = número total de observaciones.

El coeficiente etha sólo puede presentar valores positivos. Además hay que tener en cuenta que tiende a dar una asociación algo exagerada y que está influido por el número de columnas o de filas de la tabla. No se debe usar para muestras pequeñas. (Ver ejercicio 197).

EL COEFICIENTE θ DE WILCOXON

Es apropiado para medir la asociación entre una variable de escala nominal y otra de escala ordinal.

La fórmula y el procedimiento de obtener de este coeficiente se explican en el ejercicio 195. (Ver también el 189).

EJERCICIOS

188

Dadas las siguientes tablas que figuran en Goode y Hatt (1952, página 518):

AMPLITUD DE LA CASA HABITADA Y AMPLITUD DE LA FAMILIA

Amplitud de la familia	*Porcentajes de familias que residen en casas*		
	Grandes	*Medianas*	*Pequeñas*
Familias grandes	25	55	20
Familias medianas	30	60	10
Familias pequeñas	30	55	15

AMPLITUD DE LA CASA HABITADA POR FAMILIAS GRANDES, SEGUN EL NIVEL ECONOMICO DE LA FAMILIA (En porcentajes)

Nivel económico de la familia	*Porcentajes de familias grandes que viven en casas*		
	Pequeñas	*Medias*	*Grandes*
Alto	5	35	60
Medio	25	65	10
Bajo	40	55	5

Se pide:

a) Determinar en qué sentido se debe hallar épsilon.

b) Calcularlo.

c) Interpretar los resultados de las dos tablas conjuntamente en relación a la hipótesis de que la amplitud de la casa está en relación con el tamaño de la familia.

RESPUESTA

a) Epsilon se debe hallar en sentido vertical, pues los porcentajes figuran obtenidos en sentido horizontal, como se puede comprobar, viendo que suman 100 en este sentido, pero no en el contrario.

b) Cálculo. Primera tabla:

Epsilons de la columna de casas pequeñas = 5; 0 y 5.

Epsilons de la columna de casas medias = 5; – 5 y 0

Epsilons de la columna de casas grandes = 10; – 5 y 5.

Segunda tabla:

Epsilons de la columna de casas pequeñas = 20; 20 y 40.

Epsilons de la columna de casas medias = 30; – 10; 20.

Epsilons de la columna de casas grandes = – 50; – 5; – 55.

Interpretación: Según los épsilons obtenidos en la primera tabla, dado su bajo nivel en todas las columnas, no aparece una relación significativa entre el tamaño de la familia y la amplitud de la casa, lo que lleva a rechazar la hipótesis indicada tal como figura formulada.

En cambio, según los épsilons de la segunda tabla, sí hay una relación positiva entre el nivel económico y el tanto por ciento de familias grandes que residen en casas grandes. Se ve que cuanto mayor es el nivel económico de la familia mayor es el porcentaje de familias que residen en casas grandes, y al revés en las pequeñas. Por tanto, la hipótesis se podría modificar de acuerdo con los resultados de esta segunda tabla, diciendo que la amplitud de la casa se halla en relación con el tamaño de la familia si se controla el nivel de renta o si el nivel de rentas de las familias consideradas es el mismo.

189

Dadas las siguientes TABLAS:

I. RENTA PER CAPITA (1973) y GASTOS DE CONSUMO POR HOGAR (1974) en algunas REGIONES ESPAÑOLAS, en MILES DE PESETAS (redondeadas).

Regiones	*Renta*	*Gasto consumo*
Andalucía	72	220
Aragón	98	260
Canarias	83	300
Cataluña	127	310
Extremadura	60	180
Castilla la Vieja	97	250
Galicia	73	220
Navarra	109	300

Fuentes: Banco de Bilbao: Renta nacional de España y su distribución provincial.
INE: Encuesta de presupuestos familiares: Madrid, 1975.

II. GRADO DE FELICIDAD SEGUN LA OCUPACION EN ESPAÑA, 1970, en números absolutos.

GRADO DE FELICIDAD	*OCUPACION*			
	Universitarios	*Profesionales*	*Obreros y empleados*	*TOTAL*
Muy feliz	32	67	78	177
Bastante feliz	141	121	248	510
Poco feliz	59	27	74	160
Nada feliz	12	9	12	33
N	244	224	412	880

Fuente: Resumen del informe Foessa. España. Euramérica, 1970. p. 179

III. DESEO DE TENER MAYOR PARTICIPACION EN LA VIDA POLITICA DE ESPAÑA SEGUN EL NIVEL DE ESTUDIOS; En España, 1975

DESEO DE PARTICIPAR EN LA VIDA POLITICA	*ESTUDIOS*			
	Ninguno y primarios	*Medios*	*Superiores*	*TOTAL*
Mucho y bastante	280	180	100	560
Ni poco ni mucho	320	100	30	450
Nada	1110	230	50	1390

N	1710	510	180	2400

Fuente: Salustiano del Campo y otros. La cuestión regional española. Madrid. Edicusa.

1) Enunciar las hipótesis que suponen las tablas.

RESPUESTA

Toda tabla de más de una variable supone implícitamente la hipótesis de que las variables que comprende están asociadas en su variación.

Hipótesis que implican las tablas en este caso:

a) A mayor nivel de renta "per cápita", mayor nivel de gastos de consumo por hogar.

b) Existe una correlación entre ocupación y grado de felicidad. El grado de felicidad es diferente en las distintas ocupaciones.

c) A mayor nivel de estudios, mayor deseo de participar en la vida política.

2) Señalar la clase de tabla y el tipo de escala de las variables que forman las distribuciones.

RESPUESTA

Tabla 1ª. Clase de tabla: Distribución bivariante conjunta.

Variables: Renta per cápita y gasto de consumo por hogar. Las dos son variables de razón o proporción, porque tiene distinción, orden, unidad de medida, la peseta, y origen cero.

Tablas II y III. Las dos son de contingencia.

Variables de la tabla II: Ocupación y grado de felicidad; nominal la primera y ordinal la segunda.

Tabla III. Nivel de estudios y deseo de participar en la vida política; las dos de escala ordinal.

3) Calcular los coefificientes gamma, lambda, rho y zeta de Wilcoxon, previa indicación de los que correspondan en cada tabla, permitiéndose considerar a este respecto la variable grado de felicidad como nominal. Interpretar los resultados.

RESPUESTA

Los coeficientes que corresponde hallar en cada tabla son:

En la primera, el coeficiente rho, válido para tablas bivariantes conjuntas con variables de intervalo o de razón cuando el número de casos es inferior a 30.

En la segunda, el coeficiente zeta de Wilcoxon, por tratarse de una variable nominal y otra ordinal. También es aplicable el coeficiente lambda si se considera, como se permite en el enunciado, la variable grado de felicidad como nominal, por tratarse entonces de dos variables nominales.

En la tabla tercera, el coeficiente gamma para tablas de contingencia por ser las dos variables ordinales.

Cálculo del coeficiente rho.

Disposición de los cálculos:

Variables		*Rangos*		*d*	d^2
Renta	*Consumo*	*Renta*	*Consumo*		
72	220	7	6,5	0,5	0,25
98	260	3	4	1	1
83	300	5	2,5	2,5	6,25
127	310	1	1	0	0
60	180	8	8	0	0
97	250	4	5	1	1,00
73	220	6	6,5	0,5	0,25
103	300	2	2,5	0,5	0,25
					9,00

$$\rho = 1 - \frac{6\,(\Sigma\, d^2)}{N\,(N^2-1)} = 1 - \frac{6\,(9)}{8\,(8^2-1)} = 1 - \frac{54}{504} = 1 - 0{,}13 = 0{,}87$$

Interpretación: El alto coeficiente obtenido revela una fuerte correlación entre las variables de la tabla que confirma la hipótesis indicada.

Coeficiente lambda.

Fórmula:

$$\frac{\Sigma fc - Fh}{N - Fh} = \frac{(141 + 121 + 248) - 510}{870 - 510} = \frac{0}{360} = 0$$

Coeficiente zeta de Wilcoxon.

Fórmula:

$$\theta = \frac{\Sigma\, Di}{T^2}$$

Comparación entre universitarios y profesionales.

$$32\,(121 + 27 + 9) + 141\,(27 + 9) + 59\,(9) = 10.631$$

De abajo a arriba.

$$12\,(27 + 121 + 67) + 159\,(121 + 67) + 141\,(67) = 23.119$$

Diferencia:

12.488

Comparación entre universitarios y obreros:

$$32\,(248 + 74 + 12) + 141\,(74 + 12) + 59\,(12) = 23.522$$

$$12\,(74 + 248 + 78) + 59\,(248 + 78) + 141\,(78) = 38.338$$

Diferencia 14.816

Comparación entre profesionales y obreros:

$$67\,(248 + 74 + 12) + 121\,(74 + 12) + 27\,(12) = 37.206$$

$$9\,(74 + 248 + 78) + 27\,(248 + 78) + 121\,(78) = 21.840$$

Diferencia 15.366

Sumatorio Diferencias:

$$12.488 + 14.816 + 15.366 = 42.670$$

$$T^2 = 244\,(224) + 244\,(412) + 224\,(412) = 247.472$$

$$\theta = 42.670/247.472 = 0{,}17$$

Interpretación: El coeficiente zeta obtenido indica una correlación muy baja y el lambda la no asociación o independencia entre las variables de la tabla. La diferencia entre ambos resultados se explica porque, al considerar en lambda la variable ordinal como nominal, se pierde exactitud.

Coeficiente gamma:

Fórmula:

$$\frac{fc - fi}{fc + fi}$$

$$fc = 280\,(100 + 30 + 230 + 50) + 320\,(230 + 60) = 204.400$$

$$fi = 1110\,(100 + 30 + 180 + 100) + 320\,(180 + 100) = 544.700$$

$$Y = \frac{204.400 - 544.700}{204.400 + 544.700} = \frac{343.700}{749.100} = -0{,}45$$

El coeficiente obtenido señala una correlación negativa apreciable entre el nivel de estudios y la preocupación política, que contradice la hipótesis de partida y el

sentido común. La explicación de esta anomalía se encuentra en la tabla, que está mal construida, pues el orden de las categorías de la variable estudios va de menos a más y el de la variable preocupación política de más a menos. Si se colocaran ambas variables en el mismo orden, el signo del coeficiente sería positivo y resultaría confirmada la hipótesis.

190

En un estudio sociológico sobre la emigración española laboral a distintos países de Europa se ha formado con los resultados obtenidos la tabla que se indica a continuación:

	Emigración (en miles) (b)				
Regiones (a)	*Francia*	*Suiza*	*Alemania*	*Bélgica*	*Total*
Ambas Castillas	175	80	125	60	440
Andalucía y Extremadura	150	50	220	80	500
Galicia	180	100	200	40	520
Asturias, Vascongadas y Navarra	225	75	100	50	450
Aragón y Murcia	170	80	130	20	400
Cataluña, Valencia y las Islas	250	50	50	25	375
Suma	1.150	435	825	275	2.685

Indicar, primero, el tipo de escala de las variables de la tabla, y hallar el coeficiente de correlación lambda de los datos de la tabla, considerando primero como variable independiente la emigración, después las regiones y, por último, en conjunto sin tener en cuenta una variable independiente determinada.

RESPUESTA

El tipo de escala de las dos variables es nominal. Las categorías que las forman sólo implican distinción, pero no orden, distancia ni origen.

El coeficiente aplicable es, por tanto el lambda.

Sustituyendo en la fórmula expuesta en la parte teórica:

$$\lambda_{ba} = \frac{(250 + 100 + 220 + 80) - 520}{2.685 - 520} = \frac{650 - 520}{2.165} = 0{,}06$$

El coeficiente obtenido, 0,06 ó 6 por 100, es muy bajo. Indica que el conocimiento del país de emigración sólo elimina un 6 por 100 de los errores cometidos al predecir la emigración según regiones.

$$\lambda_{ab} = \frac{(175 + 220 + 200 + 225 + 170 + 250) - 1.150}{2.685 - 1.150} = \frac{90}{1.535} = 0{,}05$$

o el 5 por 100. Resultado también muy bajo. El conocimiento de la región sólo elimina el 5 por 100 de los errores al predecir el país de destino de la emigración. Solución sin determinar la variable independiente:

$$\lambda = \frac{650 + 1.240 - 1.150 - 520}{2 \times 2.685 - (520 + 1.150)} = \frac{1.890 - 1.670}{5.370 - 1.670} =$$

$$= \frac{220}{3.700} = 0{,}06;\ \text{ó}\ 6\ \%$$

Coeficiente también muy bajo. Indica en general, como los anteriores, que no hay apenas relación entre la región de origen y el país de destino de la emigración.

191

En una investigación sociológica sobre una pequeña población y para contrastar la hipótesis de que la fecundidad familiar disminuye con el aumento del nivel de estudios, se ha obtenido la siguiente tabla estadística que relaciona las variables alfabetización y fecundidad familiar.

Hallar e interpretar la correlación entre ambas mediante el coeficiente phi.

	Fecundidad familiar	
Alfabetización	*Alta (más de cuatro hijos por matrimonio*	*Baja (menos de dos hijos por matrimonio*
Analfabetos	140 (a)	70 (b)
Alfabetizados	30 (c)	120 (d)
Total	170	190

RESPUESTA

Fórmula a aplicar:

$$\frac{a \cdot d - d \cdot c}{\sqrt{(a+b) \cdot (c+d) \cdot (a+c) \cdot (b+d)}}$$

$$\frac{(120 \times 140) - (30 \times 70)}{\sqrt{(140+70) \times (30+120) \times (140+30) \times (70+120)}} =$$

$$= \frac{16.800 - 2.100}{\sqrt{210 \times 150 \times 170 \times 190}} = \frac{14.700}{31.897} = 0{,}46$$

El coeficiente obtenido indica una asociación importante positiva entre ambas variables.

Nos dice que a mayor alfabetización mayor fecundidad familiar. El signo tiene significación en ese caso porque hay orden en las categorías de las variables. Sin embargo este resultado contradice la hipótesis a contrastar que postula una correlación negativa y no positiva entre ambas variables. Ello obliga a revisar detenidamente la tabla y las operaciones efectuadas. Esta revisión muestra que la tabla está mal construída. Las dos categorías de las dos variables presenta un orden. Sin embargo, este orden está dispuesto en la tabla inversamente en cada variable, de más a menos en la fecundidad (alta-baja) y de menos a más (analfabetos-alfabetizados) en la alfabetización. Esta es, pues, la razón o el motivo de que el signo del coeficiente resulte positivo y no negativo como corresponde según los datos de la investigación y se puede comprobar repetiendo el ejercicio previa variación del orden en la tabla de las categorías de la variable alfabetización, por ejemplo.

192

En una investigación comparativa entre gitanos y el resto de la población, en una pequeña comunidad del sur de España se han obtenido los siguientes resultados:

Nivel de estudios	*No gitanos*	*Gitanos*
Secundario	5	0
Primario	225	45
Ninguno	100	95

Hallar e interpretar la correlación entre ambas variables por medio del coeficiente Q de Yule:

1. Conforme está.

2. Considerando en el nivel de estudios los grados secundario y primario.

3. Considerando los grados primario y ninguno.

RESPUESTA

Hay que aplicar la fórmula indicada en la parte teórica. Sin embargo, a simple vista se advierte que esta fórmula no puede aplicarse en este caso porque no se trata de una tabla de contingencia de dos por dos, o sea, de cuatro datos, sino de dos por tres, en la que a, b, c y d no abarcan los seis datos.

Si se consideran sólo los grados secundario y primario en el nivel de estudios, tampoco se puede aplicar, pues al ser el primero de ellos 0, la fórmula nos daría necesariamente como resultado 0, cualquiera que sean las demás cantidades.

En cambio, sí se puede aplicar si se reducen los grados de nivel de estudios a primario y ninguno.

La tabla de contingencia sería en este caso:

ESTUDIOS	*No gitanos*	*Gitanos*
Primarios	230	45
Ninguno	100	95

Aplicando la fórmula:

$$Q = \frac{95 \times 225 - 45 \times 100}{45 \times 100 + 45 \times 225} = Q = \frac{21.850 - 4.500}{4.500 + 21.850} = \frac{17.350}{26.350}$$

$$Q = 0{,}65$$

El coeficiente obtenido indica una asociación importante entre las dos variables y la probabilidad de acierto en la predicción de una, conociendo la otra.

193

En una investigación se ordenó primero por una muestra de padres obreros las cualidades de los hijos, según su importancia para su formación. Después se pidió a una muestra de padres profesionales que ordenaran nuevamente según su criterio dichas cualidades. Los resultados fueron los siguientes:

	Rango de padres obreros	*Rango de profesionales*
1. Que sean honestos	12	11
2. Que obedezcan a sus padres	11	12
3. Que se comporten seriamente	10	3
4. Que sean capaces de controlarse	9	2
5. Que se pueda contar con su palabra	8	1
6. Que tengan consideración con los demás	7	4
7. Que sea afectuoso	6	6
8. Limpieza y buena presencia	5	7
9. Que sea buen estudiante	4	10
10. Que tenga espíritu investigador	3	9
11. Que sepa defenderse	2	5
12. Que tenga ambiciones de triunfar	1	8

Se pide establecer la correlación entre ambas ordenaciones, mediante la γ para rangos, e interpretarla.

La fórmula a aplicar se encuentra en el ejercicio 189.

RESPUESTA

Item	*Concordancias*	*Inversiones*
12	4	7
11	6	4
10	3	6
9	2	6
8	2	5
7	2	4
6	2	3
5	4	0
4	3	0
3	2	0
2	0	1
1	0	0
	30	36

Aplicando la fórmula se tiene:

$$\gamma = \frac{30 - 36}{30 + 36} = \frac{-6}{66} = -0{,}09$$

El resultado obtenido indica una correlación baja entre ambas ordenaciones de sentido negativo. Existe una oposición de padres obreros y profesionales sobre las cualidades importantes en sus hijos, pero es despreciable o no significativa.

194

Según el "Anuario Estadístico de España", edición manual de 1971, página 510, los índices de nupcialidad y mortalidad de las diez primeras provincias españolas, por orden alfabético, del año 1970, fueron los que se indican en la tabla siguiente. Se pide calcular la correlación entre ambas series de índices utilizando la técnica llamada coeficiente Rho de Spearman.

Provincias	*Tasa de nupcialidad por 1.000 habitantes*	*Tasa de mortalidad por 1.000 habitantes*
Alava	7,24	7,65
Albacete	7,67	10,08
Alicante	7,75	9,36
Almería	7,68	8,78
Avila	6,78	9,73
Badajoz	6,81	8,93
Baleares	7,54	11,34
Barcelona	7,77	8,02
Burgos	6,80	9,03
Cáceres	7,66	8,62

Interpretar el resultado que se obtenga y razonar la aplicación de este coeficiente.

RESPUESTA

Aunque las dos variables de la tabla tasa de nupcialidad y tasa de mortalidad lo son de escala de razón y, en consecuencia, el coeficiente apropiado sería el r de Pearson, sin embargo se admite la aplicación del rho de Spearman, al ser el número de casos menor de 30.

			Grados			
PROVINCIAS	*Tasa nupcialidad*	*Tasa mortalidad*	*Nupcialidad*	*Mortalidad*	*Diferencia entre grado*	*Diferencia al cuadrado*
Alava	7.24	7,65	7	10	—3	9
Albacete	7.67	10,00	4	2	2	4
Alicante	7,75	9,36	2	4	—2	4
Almería	7,68	8,78	3	7	—4	16
Avila	6,78	9,73	10	3	7	49
Badajoz	6,81	8,93	8	6	2	4
Baleares	7,54	11,34	6	1	5	25
Barcelona	7,77	8,02	1	9	—8	64
Burgos	6,80	9,03	9	5	4	16
Cáceres	7,66	8,62	5	8	—3	9
					Σd^2	200

Aplicándo la fórmula:

$$\rho = 1 - \frac{6\,(\Sigma\, d^2)}{N \times (N^2 - 1)}$$

se tiene:

$$\rho = 1 - \frac{6 \times 200}{10 \times (10^2 - 1)} = 1 - \frac{1.200}{990} = 1 - 1{,}21 = -0{,}21$$

El coeficiente de correlación Rho es bajo y negativo, lo que quiere decir que hay asociación inversa de signo negativo entre las variables nupcialidad y mortalidad referidas a las provincias indicadas.

195

En un estudio de sociología política se ha puesto en relación la filiación política y el nivel de estudios. Determinar la asociación entre estas dos variables empleando el coeficiente θ de Wilconson, para relación entre escalas ordinales y nominales, teniendo en cuenta que:

1. Se compara cada clase de escala nominal (filiación política en este caso) con cada una de las demás clases de la misma escala, es decir, de dos en dos.

2. En esta comparación cada uno de los valores de la primera clase comparada se multiplican primero por los de la otra clase con la que se le compara que estén a la derecha de dicho valor y se suman, y, después con los que estén a su izquierda y también se suman.

3. Se obtienen las diferencias entre las sumas de los productos hacia la derecha y hacia la izquierda de cada clase, y se suman todas estas diferencias sin tener en cuenta el signo.

4. Por último se aplica la fórmula:

$$\theta = \frac{\Sigma Di}{T^2}$$

en la que Di es la suma de las diferencias indicadas y T^2 el número total de comparaciones posibles, que se obtiene multiplicando el total de cada clase por los de las demás y sumando los productos.

	NIVEL DE ESTUDIOS				
Filiación Política	*Sin estudios*	*Primaria*	*Secundaria*	*Superior*	*Total*
Socialistas	10	25	75	40	150
Monárquicos	20	35	90	60	205
Republicanos	90	70	60	80	300

RESPUESTA

Comparaciones socialistas con monárquicos.

Frecuencias dirección derecha, fd.

$$fd. = 10\,(35 + 90 + 60) + 25\,(90 + 60) + 76 \times 60 = 10.100$$

Frecuencias dirección hacia la izquierda, fi.

$$fi. = 25 \times 20 + 75\,(20 + 35) + 40\,(20 + 25 + 90) = 10.425$$

Diferencia comparación socialistas con monárquicos:

$$10.100 - 10.425 = -325$$

Comparación de socialistas con republicanos:

$$fd. = 10\,(70 + 60 + 80) + 25\,(60 + 80) + 75 \times 80 = 11.600$$
$$fi. = 25 \times 90 + (90 + 70) + 40\,(90 + 70 + 60) = 23.050$$

Diferencia socialistas con republicanos:

$$11.600 - 23.050 = -11.450$$

Comparaciones monárquicos con republicanos:

$$fd. = 20\,(70 + 60 + 80) + 35\,(60 + 80) + 90 \times 80 = 16.300$$
$$fi. = 35\,(90) + 90\,(90 + 70) + 60\,(90 + 70 + 60) = 30.750$$

Diferencia monárquicos con republicanos:

$$16.300 - 30.750 = -14.450$$

Suma de las diferencias:

$$325 + 11.450 + 14.450 = 26.225$$

Número total de comparaciones: T_2

$$T_2 = 150 \times 205 + 205 \times 300 + 150 \times 300 = 30.750 + 45.000 + 61.500 = 137.250$$

Aplicando la fórmula se tiene:

$$\frac{26.225}{137.250} = 0{,}19$$

Interpretación: El coeficiente obtenido es bajo. La base para predecir el nivel de estudios basándose en la filiación política es baja. El obtenido nos dice que sólo un 19 por 100 de las comparaciones efectuadas entre personas clasificadas en distinta clase según la filiación política presenta diferencias sistemáticas en el nivel de estudios.

196

En una investigación sociológica han resultado los siguientes valores de las variables de intervalos que se indican:

X, número de faltas de asistencia a clase en el trimestre: 5, 8, 10, 15, 17, 20; e Y, media de las puntuaciones obtenidas: 12, 7, 5, 3, 2, 1;

Se pide hallar el coeficiente de correlación de Pearson entre ambas, interpretarlo e indicar su aplicación, teniendo en cuenta que la fórmula pertinente es la siguiente:

$$r = \frac{\Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{N}}{\sqrt{\left[\Sigma X^2 - \frac{(\Sigma X)^2}{N}\right] \times \left[\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}\right]}}$$

RESPUESTA

Disposición de los cálculos:

X	Y	X^2	Y^2	XY
5	12	25	144	60
8	7	64	49	56
10	5	100	25	50
15	3	225	9	45
16	2	256	4	32
20	1	400	1	20
74	30	1.070	232	263

$$r = \frac{263 - \frac{74 \times 30}{6}}{\sqrt{\left(1.070 - \frac{74^2}{6}\right) \times \left(232 - \frac{30^2}{6}\right)}}$$

$$r = \frac{263 - 370}{\sqrt{1.070 - 912{,}66) \times (232 - 150)}} = \frac{-107}{\sqrt{12.901{,}88}} = \frac{-107}{113{,}58} = -0{,}94$$

El coeficiente r obtenido, – 0,94, indica que existe una elevada asociación negativa entre las variables. Con base en él se puede predecir con gran probabilidad de acierto que al variar una de ellas en un sentido, la otra cambiará en proporción similar en en dirección contraria.

197

En una investigación sociológica sobre la circulación automovilista se estudió la relación del número de las sanciones sufridas por infracción del Código de Circulación con la siguiente tipología del carácter social de una muestra de conductores: introvertido, extrovertido, equilibrado.

Se pide hallar la correlación entre ambas variables que figuran en la tabla adjunta por el procedimiento del coeficiente etha.

Indicar el tipo de las variables.

Numero de sanciones sufridas	*Número de conductores sancionados*	*Carácter*		
		Extrovertido	*Introvertido*	*Equilibrado*
1	10	0	0	10
2	40	10	10	20
3	90	30	40	20
4	120	60	50	10
5	80	50	25	5
6	60	50	10	0
	400			

Fórmula a aplicar:

$$\eta^2 = \frac{\sum_{j=1}^{k} n_j (\overline{Y}_j - \overline{Y})^2}{\sum (Y_i - \overline{Y})^2}$$

n_j = *Número de observaciones en el subgrupo j de X.*
$\overline{Y}$ = *Media aritmética general.*
$\overline{Y}_j$ = *Media del subgrupo j.*
K = Número de subgrupos.
Y_i = *Valor de la escala de intervalos.*
N = *Número total de observaciones.*

RESPUESTA

Y	f	f × Y	Y	$Y - \bar{Y}$	$(Y - \bar{Y})^2$	f. $(Y - \bar{Y})^2$	Extr.	f. Y	Intr.	f. Y	Equi.	f.Y
1	10	10	4	—3	9	90	0	0	0	0	10	10
2	40	80	4	—2	4	160	10	20	10	20	20	40
3	90	270	4	—1	1	90	30	90	40	120	20	60
4	120	480	4	0	0	0	60	240	50	200	10	40
5	80	400	4	1	1	80	50	250	25	125	5	25
6	60	360	4	2	4	240	50	300	10	60	0	0
	400	1.640			19	660	100	900	135	525	65	175

$\bar{Y}$ extrovertidos = 900 : 200 = 4,50 − 4 = $0,50^2$ = 0,25

Y introvertidos = 525 : 135 = 3,88 − 4 = $-0,12^2$ = 0,01

Y equilibrados = 175 : 65 = 2,69 − 4 = $1,31^2$ = 1,71

$$\text{Etha}^2 = \frac{200 \times 0,25 + 135 \times 0,01 + 65 \times 1,71}{660} = \frac{168,50}{660} = 0,24$$

$$\text{Etha} = 0,48$$

Interpretación: Un 48 por 100 de la varianza en la frecuencia de las infracciones está asociada con el carácter social del conductor. La correlación es estimable. El conocer los tipos caracteriales puede tener significación práctica para predecir la frecuencia de las infracciones.

La variable, número de sanciones sufridas, es de tipos de escala de razón o proporción y, por tanto, también de intervalo. La variable, carácter social de los conductores, se puede considerar como tipo de escala nominal.

26. OTROS COEFICIENTES

COEFICENTES DE CORRELACION BISERIAL Y BISERIAL PUNTUAL

Naturaleza.— Estos coeficientes constituyen fundamentalmente una aplicación del coeficiente de correlación r de Pearson.

Aplicaciones.— Se utilizan para hallar la covariación entre una variable dicotómica y otra de intervalo o de razón.

La diferencia entre ellos consiste en que, en el coeficiente punto biseral, la variable dicotómica lo es realmente, mientras que, en el biseral, la dicotomía es producida por el investigador y en realidad oculta una variedad de valores o puntuaciones.

Requisitos.— Se fundan en el supuesto de una distribución normal de las variables en la población. Por tanto, si el supuesto de que se parte es la anormalidad de dichas variables en la población investigada, no deben usarse estos coeficientes.

Propiedades.— El punto biserial varía de + 1 a −1. El biserial puede alcanzar valores inferiores a −1 y superiores a +1, lo que indica que las puntuaciones de la variable no se distribuyen normalmente.

Fórmulas:

Coeficiente punto biserial:

$$r_{pb} = \frac{\bar{X}_{.1} - \bar{X}_{.0}}{\sigma_X} \sqrt{\frac{n_1 \; n_0}{n\,(n-1)}}$$

donde: $\bar{X}_{.1}$ es la media en X de las puntuaciones 1 en Y;
$\bar{X}_{.0}$ es la media en X de las puntuaciones 0 en Y;
σ_X es la desviación típica de X;
n_1 es el número de las puntuaciones con valor 1 en Y;
n_0 es el número de puntuaciones con valor 0 en Y;

$$n = n_1 + n_0$$

Coeficiente biserial:

$$r_b = \frac{X_{.1} - X_{.0}}{\sigma_X} \cdot \frac{n_1\ n_0}{u\ n\sqrt{n^2 - n}}$$

donde las equivalencias de los signos son las antes indicadas, excepto u, que es la ordenada (es decir la altura) de la curva normal que corresponde a un área de esta curva igual a n_1/n. (Se encuentra en las tablas.) (Ver ejercicios 198 y 199)

COEFICIENTE DE CORRELACION BISERIAL PARA RANGOS O POSICIONES

Naturaleza.– Es un coeficiente no paramétrico, no basado en los conceptos de media y varianza, sino en los de acuerdos e inversiones.

Aplicación.– Es apropiado para hallar la correlación entre una variable dicotómica X y una variable ordinal Y.

Requisitos.– No exige la distibución normal en la población de las variables.

Fórmula:

$$r_{br} = \frac{A - I}{n_0\ n_1}$$

en la que A son los acuerdos en las ordenaciones;
I son las inversiones;
n_0, el número de puntuaciones con valor 0 en X;
n_1, el número de puntuaciones con valor 1 en X.

Cómputo.– Para computar los acuerdos e inversiones se ordena por rangos de mayor a menor la variable ordinal Y. Después se separan los rangos que correspondan a un valor 1 de la variable dicotómica de los que correspondan a un valor cero de la misma variable.

Son acuerdos los rangos que aparecen en la columna cero inferiores a cada uno de los rangos de la columna 1.

Inversiones son los rangos que aparecen en la columna 1 inferiores a cada uno de los rangos de la columna cero.

Existe una fórmula que evita calcular así los rangos. Es la siguiente:

$$r_{br} = \frac{2}{n}(\bar{Y}_1 - \bar{Y}_0)$$

donde Y_1 es el rango promedio de las puntuaciones 1 en X y Y_0 el rango promedio de las puntuaciones cero en X.
Ver ejercicio 200.

PRUEBA DE SIGNIFICACION

Coeficiente biserial.– El error típico de este coeficiente cuando ni n_0/n ni n_1/n son menores que .05 tiene la siguiente fórmula:

$$\sigma_{rb} = \sqrt{\frac{(n_{.0}/n)\,(n_{.1}/n)}{u^2\ n \cdot \sqrt{N}}}$$

Si el r_b obtenido es superior en cinco veces al error típico se concluye que dicho coeficiente no puede proceder casualmente de una población donde la correlación entre las variables del supuesto sea cero.

Coeficiente punto biserial.– Según Guildford (1973, p. 299) como este coeficiente depende de la diferencia entre las medias, $\bar{X}_{.1}$, $\bar{X}_{.0}$ puede utilizarse para probar su significación el test de la diferencia entre medias.

También puede emplearse para contestar si este coeficiente es distinto o no de cero, la siguiente fórmula:

$$t = r_{pb_i}\sqrt{\frac{N - 2}{1 - r^2_{pb_i}}}$$

(Ver ejercicios 198 a 200)

COEFICIENTE DE CORRELACION W DE KENDALL

Naturaleza.– Este coeficiente es de naturaleza no paramétrica. No se basa en la media ni en la desviación típica. Constituye una extensión de los coeficientes para hallar la asociación entre dos ordenaciones de rangos.

Aplicaciones.– Tiene aplicación cuando interesa calcular el grado de correlación o concordancia no entre dos variables, como el coeficiente gamma de Goodman y Kruskal, sino entre más de dos transformadas en ordenaciones por rangos.

Propiedades.– Varía de +1 a –1.

Requisistos.— Sólo exige que las variables alcancen un nivel de medida ordinal y que sean más de dos.

Fórmulas.— Se distiguen dos casos, según haya o no empates en las puntuaciones de las variables.

En el caso de que no haya empates la fórmula es:

$$w = \frac{S}{\frac{1}{12} k^2 (N^3 - N)} ; \qquad S = \Sigma R_j - \frac{\Sigma R_j}{N}^2$$

donde: k = Número de conjuntos de rangos o de ordenaciones.
N = Número de individuos o de cosas ordenados.
$k^2 (N^3 - N)$ = Máxima suma posible de desviaciones al cuadrado.
R_j = suma de los rangos en cada variable que corresponda a cada unidad de observación.

En caso de empates, si éstos son pocos, se puede usar la fórmula anterior porque la diferencia es insignificante.

Si el número de empates es elevado, se puede introducir en la fórmula general el siguiente factor de corrección:

$$T = \Sigma \frac{t^3 - t}{12}$$

donde t = número de observaciones en un grupo ligadas por el mismo rango;
Σ = Sumar todas las ligas dentro de cada rango.

Con el factor de corrección incorporado, la fórmula es la siguiente:

$$w = \frac{S}{\frac{1}{12} K^2 (N^3 - N) - K \Sigma T}$$

Prueba de significación.— Cuando en este coeficiente N, el número de unidades de observación es mayor que 7, se puede contrastar si el coeficiente w de concordancia es significativamente distinto de cero, mediante la distribución muestral de ji cuadrado, aplicando la siguiente fórmula para N – 1 grados de libertad.

$$\chi^2 = k (N - 1) w$$

Si N < 7 y k o el número de grupos o variables que se comparan varían de 3 a 20 se puede utilizar la tabla de la pág. 609 de los valores críticos para el nivel de significación del 5 y el 1% en términos del s de la fórmula del coeficiente w.

(Ver ejercicios 201 y 202)

COEFICIENTE DE CONSISTENCIA K

Naturaleza.— Cuando se tienen que ordenar por rangos objetos, cosas, personas, etc, se puede pedir a los jueces o personas que tienen que hacer la ordenación, que, en lugar de realizar una ordenación directa de todos los objetos elijan un objeto o persona dentro de cada uno de todos los pares posibles que formen los objetos o personas a comparar.

Este procedimiento se llama de comparaciones pareadas y se basa en él uno de los procedimientos que existen para valorar las escalas experimentalmente.

Aplicaciones.— Dado que en estas comparaciones pareadas puede haber elecciones inconsistentes, en el sentido de que si un juez prefiere A a B y B a C, puede preferir C a A en lugar de A a C como sería consistente con sus anteriores elecciones, el coeficiente k se utiliza para medir el grado de consistencia en el conjunto de elecciones hechas por un individuo o por un grupo.

Propiedades.— Este coeficiente varía de 0 a 1. El 0 indica la máxima inconsitencia y el 1 la consistencia perfecta.

Fórmulas.— Son distintas según sea N par o impar.

$$\text{N Impar: } K = \frac{S_R^2 \cdot 12}{N^2 - 1}$$

$$\text{N Par: } K = \frac{S_R^2 \cdot 12}{N^2 - 4}$$

donde R es, en la matriz cuadrada de preferencias por pares de unos objetos respecto de otros, la suma de dichas preferencias por filas, y S_R^2 es la varianza de la columna que forman el conjunto de sumas de las hileras.

Interpretación.— El K obtenida indica la proporción o el grado de consistencia existente en relación al total de comparaciones. P. ej. si es .40 indica que el 40 % de las comparaciones son consistentes y el 60% inconsistentes.
(Ver ejercicio 203)

EJERCICIOS

198

Dada la siguiente tabla:

Unidad de Observación	*Estado Civil Casado = 1; Soltero = 0* X	*Renta mensual en miles de ptas.* Y
A	1	35
B	0	20
C	1	48
D	1	65
E	0	25
F	1	38
G	0	19
H	1	72
I	0	10

A) Determinar las variables de la tabla y el coeficiente que sería apropiado para hallar la correlación entre ellas.

RESPUESTA

La variable estado civil, se presenta, con solo dos categorías, por tanto es una variable dicotómica.

En cuanto a la variable renta, es una variable cuantitativa con distancias iguales y su origen cero; por tanto es una variable de intervalo y también de razón o proporción.

En consecuencia el coeficiente de asociación apropiado a estas variables es el biserial puntual, r.

B) Calcular este coeficiente e interpretar los resultados previa discusión de si se dan los requisitos para aplicarlo.

RESPUESTA

Sí se dan los requisitos en cuanto que la variable renta se distribuye normalmente en la población y las categorías de la variable dicotómica no encubren una continuidad subyacente.

Fórmula: $r_{pb} = \frac{X_{.1} - X_{.0}}{\sigma_X} \sqrt{\frac{n_1 n_0}{n(n-1)}}$

Para aplicar la fórmula es conveniente aplicar la siguiente tabla:

Y	X	$X_{.1}$	$X_{.0}$	X^2
1	35	35		1225
0	20		20	400
1	48	48		2304
1	65	65		4335
0	25		25	625
1	38	38		1444
0	19		19	261
1	72	72		5184
0	10		10	100
	332	258	74	15868

$$\bar{X}_{.1} = 258/5 = 51.6$$

$$\bar{X}_{.2} = 74/4 = 18.5$$

$$\sigma_X = \sqrt{\frac{15868}{9} - \left(\frac{332}{9}\right)^2} = 20.06$$

$$r_{pb} = \frac{51.6 - 18.5}{20.06} \sqrt{\frac{(5)\,4}{(9)\,8}} = .86$$

Existe una correlación importante entre el estado civil, casado-soltero, y la renta.

C) Formular la hipótesis nula y contrastar si el coeficiente hallado es distinto de cero.

RESPUESTA

Hipótesis nula: la asociación entre estado civil y la renta no es diferente de cero. Ambas variables son independientes, no están relacionadas.

Contraste:

Fórmula: $t = r_{pb} \sqrt{\frac{N-2}{1-r_{pb}^2}}$ $\qquad t = .86 \sqrt{\frac{9-2}{1-.74}} = 3.51$

Como sólo el estado civil puede ser la variable independiente se puede utilizar el test de una sola cola. En este caso, el t crítico a un nivel de significación del 5% es 1.86. Como el t observado es mayor se puede rechazar la hipótesis nula. El coeficiente obtenido es significativamente distinto de cero, a dicho nivel de significación o de probabilidad.

199

Dada la siguiente tabla:

Unidad de observación	*Calificación Aprobado = 1 Suspenso = 0* Y	*Meses de preparación* X
A	1	15
B	0	6
C	1	8
D	0	9
E	0	10
F	1	15
G	1	6
H	0	11
I	1	12
J	1	16
K	0	5

A) Determinar las variables de la tabla y el coeficiente más apropiado para hallar la correlación entre ellas.

RESPUESTA

La variable Y es una variable dicotómica con sólo dos categorías, pero con la particularidad de que esta dicotomía se puede estimar como el resultado de puntuaciones diversas continuas.

En cuanto a la variable, tiempo de preparación, es una variable cuantitativa, con distancias iguales y origen natural cero y por tanto constituye una variable de intervalo y también de razón o proporción.

En consecuencia, el coeficiente apropiado para hallar la asociación entre estas variables es el biserial, r_b.

B) Hallar este coeficiente e interpretar los resultados previa discusión de si se dan los requistos para aplicarlo.

RESPUESTA

Sí se dan los supuestos de aplicación de este coeficiente, en cuanto la variable dicotómica, Y, se puede ver en realidad como el resultado de la agrupación de puntuaciones diversas que se puede suponer se distribuyen normalmente en la población, es decir, habrá en ésta muchas puntuaciones medias y pocas extremas. Distribución normal que también se puede pensar se da en la variable X, tiempo de estudio.

Fórmula: $$r_b = \frac{X_{.1} - X_{.0}}{\sigma_x} \cdot \frac{n_1\, n_0}{u\, n\sqrt{n^2 - 1}}$$

Tabla de datos:

Y	X	$X_{.1}$	$X_{.0}$	X^2
1	15	15		225
0	6		6	36
1	8	8		64
0	9		9	81
0	10		10	100
1	15	15		225
1	6	6		36
0	11		11	121
1	12	12		144
1	16	16		256
0	5		5	25
	113	72	45	1313

$$X_{.1} = 72/6 = 10.28$$

$$X_{.0} = 41/5 = 8.2$$

$$\sigma_X = \sqrt{1313/11 - (113/11)^2} = 3.72$$

$$u = 6/11 = .55 = .3956$$

Para hallar u dividimos n_1 por $n = .55$ y buscamos en la tabla de la curva normal el valor de la ordenada para un área de .55; es .3956 = u.

$$r_b = \frac{10.28 - 8.2}{3.72} \cdot \frac{6\,(5)}{(.3956)\ 11\sqrt{121 - 1}} = .35$$

Interpretación.– Es un coeficiente positivo bajo. Hay una asociación positiva más bien reducida entre los meses de preparación y las calificaciones obtenidas.

C) Formular la hipótesis nula y contrastarla.

RESPUESTA

H: La asociación entre los meses de estudio y las calificaciones no es diferente de cero: Ambas variables son independientes.

Contraste.– Error típico:

$$\sigma_{rb} = \sqrt{\frac{(n_0/n)\ (n_1/n)}{u\, n\sqrt{N}}} = \sqrt{\frac{(6/11)\ (5/11)}{(.3956)\ 11\sqrt{11}}} = .13$$

Como el coeficiente obtenido, .35 no es superior a cinco veces el error típico (.130) (.5) = .65, no se rechaza la hipótesis nula. Ambas variables son independientes. El coeficiente obtenido no es significativamente distinto de cero.

200

Dada la siguiente tabla:

Unidad de observación	*Hombre = 1 Mujer = 0 X*	*Puntuaciones en feminismo Y*
A	1	85
B	0	40
C	1	98
D	1	102
E	0	61
F	0	78
G	1	70
H	0	86
I	1	72
J	0	30

A) Determinar las variables de la tabla y el coeficiente más apropiado para hallar la correlación entre ellas.

RESPUESTA

La variable X, el sexo, es una variable dicotómica con dos categorías, hombre y mujer.

La variable Y es una variable de escala ordinal porque aunque presente valores cuantitativos, al proceder de puntuaciones de escala Likert, no se puede decir que estas puntuaciones consten de unidades de distancia iguales entre sus distintos valores. Sólo señalan un mayor o menor grado de feminismos según la puntuación obtenida.

En consecuencia el coeficiente apropiado en este caso sería el biserial para rangos.

B) Hallar este coeficiente e interpretarlo, previa discusión de si se dan las condiciones para aplicarlo.

RESPUESTA

Sí se dan las condiciones de aplicación en cuanto éstas son, simplemente, que una variable sea dicotómica y otra ordinal.

Fórmula primera:

$$r_{br} = \frac{A - I}{n_0 (n_1)}$$

Tabla de datos:

X	*Y*	*Rangos de Y*	*Rangos de Y = 1*	*Rangos de X = 0*	*Acuerdos*	*Inversiones*
1	85	7	7		4	
0	40	2		2		0
1	98	9	9		5	
1	102	10	10		5	
0	61	3		3		0
0	78	6		6		2
1	70	4	4		3	
0	86	8		8		3
1	72	5	5		3	
0	30	1		1		0
			35	20	20	5

$$r_{br} = \frac{20 - 5}{5 (5)} = .60$$

Segunda fórmula:

$$r_{br} = 2/n (\overline{Y}_1 - \overline{Y}_0) = 2/10 (7 - 4) = .60$$

Existe una correlación positiva importante entre el sexo y la puntuación en feminismo.

201-2

Dada la siguiente tabla en la que se reflejan los rangos correspondientes a diez empresas respecto a las variables, X, productividad, Y, conflictividad social, y Z, autoritarismo:

Variable	*Empresa*									
	a	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*	*j*
X	1.	4.5	2.	4.5	3.	7.5	6	9	7.5	10
Y	2.5	1.	2.5	4.5	4.5	8.	9	6.5	10.	6.5
Z	2.	1.	4.5	4.5	4.5	4.5	8.	8.	8.	10.

A) Determinar las variables de la tabla y su tipo y el coeficiente más apropiado para hallar la correlación entre ellas.

RESPUESTA

Las tres variables son de tipo ordinal en cuanto están formadas por un conjunto de rangos.

Como son más de dos, el coeficiente de correlación adecuado es el de concordancia w de Kendall.

B) Calcular este coeficiente previa discusión si se dan los presupuestos que exige.

RESPUESTA

Al ser las variables más de dos y ordinales se cumplen las condiciones requeridas para su aplicación.

Existen empates, luego la fórmula a aplicar es:

$$w = \frac{s}{(1/12)\,k^2\,(N^3 - N) - k\,\Sigma\,T}$$

Tabla para el cálculo de s:

Empresas	X	*Rangos en* Y	Z	R_j	$\Sigma \frac{R_j}{N}$	$R_j - \Sigma \frac{R_j}{N}^2$
a	1.	2.5	2.0	5.5	16.5	121
b	4.5	1	1	6.5	16.5	100
c	2.	2.5	4.5	9.	16.5	56.25
d	4.5	4.5	4.5	13.5	16.5	9.

Empresas	X	*Rangos en* Y	Z	R_j	$\Sigma \frac{R_j}{N}$	$\left(R_j - \Sigma \frac{R_j}{N}\right)^2$
e	3.	4.5	4.5	12.	16.5	20.25
f	7.5	8.	4.5	20.	16.5	12.25
g	6.	9.	8.	23	16.5	42.25
h	9.	6.5	8.	23.5	16.5	49.
i	7.5	10.	8.	25.5	16.5	81.
j	10	6.5	10	26.5	16.5	100
				165.		591.

Cálculo de los empates: En la primera variable hay dos empates de dos unidades de observación cada uno. Para cada empate las dos unidades se elevan al cubo y se resta 2, luego,

$$T_x = \frac{(t^3 - t)}{12} \qquad T_x = \frac{(2^3 - 2) + (2^3 - 2)}{12} = 1$$

En la variable Y hay tres empates de dos unidades cada uno luego,

$$T = [\,(2^3 - 2) + (2^3 - 2) + (2^3 - 2)\,]/12 = 1.5$$

En la variable Z hay también dos empates, pero uno de ellos abarca cuatro unidades y el otro tres. Por tanto:

$$T_z = \frac{(4^3 - 4) + (3^3 - 3)}{12} = \frac{84}{12} = 7$$

$$w = \frac{591}{(1/12)\,3^2\,(10^3 - 10) - 28.5} = .827$$

El coeficiente es elevado. Hay una asociación positiva muy alta entre las ordenaciones que representan las tres variables en cuestión en relación a las empresas investigadas.

C) Formular la H_0 en relación al coeficiente obtenido y contrastar su significación del 5

RESPUESTA

H_0 : Los rangos de las empresas en las tres variables no están relacionados; las tres variables son independientes.

Contraste:

$$\chi^2 = k\,(n - 1)\,w = 3\,(10 - 1)\,.827 = 22.329$$

χ^2 crítico para 10 - 1 = 9 grados de libertad y al nivel de significación del 1%, 21.87.

El χ^2 observado es mayor, luego se rechaza la H_0 con una probabilidad de error del 1% . Las variables no son dependientes. La concordancia entre sus rangos no es debida al azar.

Dada la siguiente tabla:

Nº de veces que el item de la fila es juzgado más favorable que el item de la columna por un grupo de 342 jueces.

Nº	1	2	3	4	5	6	7 1
1	171	96	75	58	44	34	20
2	246	171	150	147	123	106	68
3	267	192	171	161	150	147	106
4	284	195	181	171	154	150	120
5	298	219	192	188	171	168	137
6	309	236	195	192	174	171	147
7	321	274	236	222	205	195	171

Fuente: K. Schuessler, 1971, p. 309.

A) Indicar las variables y el coeficiente aplicable para hallar la asociación de ellas.

RESPUESTA

Se pueden considerar dos variables representadas por unas filas y columnas de la tabla una, que expresa el número de veces que el item de la fila es juzgado por los jueces más favorables que el item de la columna, y la segunda representada por las columnas que expresan las veces que el item de la columna es juzgado más favorablemente que el item de la fila. Ambas variables son ordinales.

El coeficiente adecuado es el de consistencia, k.

B) Calcular el coeficiente e interpretarlo.

RESPUESTA

Fórmula, N-par:

$$k = \frac{S_R^2 \cdot 12}{N^2 - 4}$$

Tabla de datos:

Items	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*R*	$(R - R)^2$
1		96	75	58	44	34	20	327	488601
2	246		150	147	123	106	68	840	34596
3	267	192		161	150	147	106	1023	9
4	284	195	181		154	150	120	1084	3364
5	298	219	192	188		168	137	1202	30976
6	309	236	195	192	174		147	1253	51529
7	321	274	236	212	205	195		1453	183329
								7182	792404

$$R = 7182/7 = 1026$$

$$S_R^2 = \frac{(R - \bar{R})^2}{N} = 792404/342 = 2316.97$$

$$K = \frac{2316.97\,(12)}{(342)^2 - 4} = .23$$

Interpretación. Es un coeficiente bajo. Indica que sólo un 23 % de las elecciones en las comparaciones hechas son consistentes y un 77% inconsistentes.

Ha habido poca consistencia en la labor de los jueces.

27. TEST DE HIPOTESIS

NOCION Y OPERACIONES QUE COMPRENDEN

Los resultados de las investigaciones sociales se refieren normalmente a muestras de la población investigada y no a esta población misma. Si bien se suele hacer la generalización de que dichos resultados también son válidos para el universo o población de que se trate, sin embargo se plantea la duda fundamental de si es admisible esta generalización.

Esta duda se basa en que los resultados obtenidos en la muestra pueden ser debidos, con relación a la población, al azar y no a darse de hecho en la población el valor o la asociación en cuestión.

Existe, pues, en todo resultado obtenido de una encuesta el problema de su validez respecto a toda la población, que recibe el nombre de significación estadística.

Se concreta en comprobar si los resultados son o no explicables simplemente por las fluctuaciones del error de las muestras utilizadas. Este problema da lugar a toda una rama de la estadística ya citada, la estadística inferencial.

Partes fundamentales de esta estadística son: la teoría de la estimación y la teoría de la decisión. La primera se refiere a la validez de la estimación de parámetros de la población basándose en medidas (estadísticos) obtenidas de una muestra. Proporciona intervalos de confianza.

La finalidad de la segunda es determinar con los tests si se puede rechazar y dentro de qué límites de probabilidad la llamada hipótesis nula, la cual postula que los resultados obtenidos son debidos al azar y que no existe en la población total la relación observada en la muestra.

De acuerdo con Blalock (1966, pág. 136 y ss) los test estadísticos comportan las siguientes operaciones específicas: A. Formular supuestos. B. Obtener la distribución de muestreo. C. Seleccionar el nivel de significación y la región crítica. D. Calcular el estadístico de la prueba. E. Tomar una decisión.

A. FORMULACION DE SUPUESTOS

Estos supuestos se refieren normalmente a:

1. Nivel de medición de las variables, logrado en la obtención de los datos.

2. A ciertas características de la población de la que procede la muestra y de esta misma, tales como: a) la normalidad o no de la población; b) la igualdad o desigualdad de varianzas de las poblaciones cuando estas son dos o más; c) al tipo de muestras, cuando son varias, independientes o dependientes.

3. A ciertas propiedades de la población o poblaciones inferidas de las muestras que proceden de dichas poblaciones.

El segundo grupo de supuestos, constituyen en los test de hipótesis lo que se llama modelo (Blalock, 1966, pág. 136), es decir, aquellos supuestos que se admiten o aceptan, que el investigador da por buenos, sin someterlos a prueba en el test. Por el contrario los del grupo tercero forman las hipótesis que se contrastan con el test estadístico.

Estas hipótesis son fundamentalmente dos: la hipótesis inicial de la investigación, H_1, que afirma ciertas propiedades en la población y su contraria, la hipótesis nula, H_0, que niega se den en la población tales propiedades. Esta última es la que se somete a contraste mediante el test.

B. OBTENCION DE LA DISTRIBUCION DE MUESTREO

Según se deriva de su nombre y se deduce incluso de lo expuesto hasta aquí, los test no son otra cosa que una prueba de unos resultados obtenidos con un patrón o regla estadística determinado.

Supongamos que además de la muestra con que hemos trabajado en la investigación, sacamos de la misma población todas las muestras posibles y repetimos en ellas la investigación.

Entonces obtendríamos una serie de resultados por lo general todos ellos distintos, que se distribuirían alrededor del verdadero valor del

resultado en la población, formando una curva. De ellos, los valores centrales serían muy numerosos y, por tanto, muy probables, y los valores extremos muy poco frecuentes y, por tanto, muy poco probables.

Esta curva, que se llama curva de la distribución muestral, es la que sirve de patrón o regla para contrastar los resultados obtenidos mediante las operaciones de los test de hipótesis.

Para realizar las cosas con exactitud total habría que formar en todos los casos de contraste de hipótesis esta curva. Pero como se comprende fácilmente, ello es prácticamente imposible.

Sin embargo, en su lugar, se han ido formando una serie de curvas básicas, que son las que se utilizan para hacer el contraste.

Entre estas curvas se pueden destacar la curva de la distribución normal, ya aludida y la curva de la distribución t de Student, que se utiliza en el caso de muestras pequeñas, en la que no es apropiada la distribución normal.

Blalock (1966, pág. 160), aunque señala que hay que ser cauto en el empleo de las reglas prácticas, estima que es grande siempre o casi siempre una muestra de $N > 100$, y también de $N > 50$ si se tiene evidencia empírica de que la desviación de la población en relación a la normalidad no es importante. Por el contrario, una muestra de $N < 30$ debe estimarse siempre como pequeña salvo que se esté seguro de la normalidad aproximada de la población.

C. SELECCION DE NIVEL DE SIGNIFICACION

El aceptar o rechazar una hipótesis de acuerdo con las tablas oportunas no nos da la seguridad absoluta de haber acertado.

Aquí, como en general en todos los casos, la estadística no proporciona una seguridad total, sino una probabilidad mayor o menor.

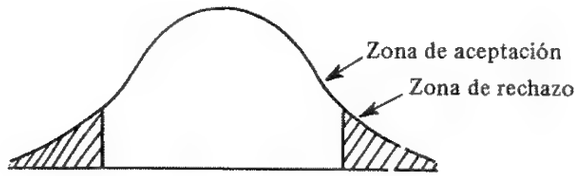
En concreto se pueden cometer dos tipos de errores que se llaman errores de tipo I, α, y errores de tipo II, β. El primer tipo consiste en rechazar la hipótesis nula siendo cierta. El segundo en aceptarla siendo falsa.

Con relación a la curva de la distribución muestral que se va a utilizar como contraste, el nivel de significación representa el señalamiento de una porción de dicha curva, en la que si caen los resultados del test estadístico dentro de ella se rechaza la Ho. Pero al hacerlo así, se corre el riesgo, como se ha indicado, de cometer el error de rechazar Ho como falsa, siendo verdadera.

En consecuencia, esta porción o nivel de significación representa la probabilidad máxima que tenemos de cometer el error de tipo I, de rechazar la hipótesis nula siendo verdadera.

El nivel de significación lo fija el investigador y depende de la opción que este haga al efecto en cada caso concreto.

Al realizar esta opción el analista divide el área que abarca la curva de la distribución muestral, utilizada como contraste en el test, en dos zonas: zona de aceptación y región crítica o zona de rechazo.



Si los resultados obtenidos en el test caen dentro de la zona de rechazo o de la región crítica, entonces son poco probables, o si se quiere, muy distintos de los normales, por tanto, se rechaza la H_0 y al contrario si caen en la zona de aceptación.

El analista debe tener en cuenta para realizar esta opción, fundamentalmente dos factores: el riesgo que está dispuesto a asumir o cometer de cada uno de los tipos indicados de errores I y II, y el tipo de prueba de una o de dos colas.

Estos errores son contrapuestos, de modo que la reducción de uno de ellos aumenta el riego de cometer el otro y viceversa.

Es obvio que si reducimos la región crítica disminuye el riesgo de error de tipo I, pero al mismo tiempo aumenta la zona de aceptación de la H_0 y con ello el riesgo si los resultados caen dentro de ella de aceptación de la H_0 siendo falsa.

Por ello, el analista debe sopesar ambos riesgos y decidir el nivel de significación de acuerdo con que, según las circunstancias del caso, le interese más: evitar un error de tipo I, caso en el que elegirá una región crítica pequeña o por el contrario le interesa evitar el error de tipo II, caso en el que elegirá una zona de aceptación de la H_0 pequeña y por tanto una región crítica grande.

En las tablas en las que se ha realizado este contraste estadístico se consigna el nivel de significación utilizado, con la letra p, por ejemplo, $p < 0{,}01$, lo que indica la significatividad de la medida estadística de que se trate a un nivel de probabilidad menor del 1 por 100.

En la práctica los niveles de significación o áreas de la región crítica, que se suelen utilizar son el 5 y el 1 por 100 de dicha área.

Estos niveles de significación, significan que se adopta la probabilidad de 5 y de 1 sobre cien respectivamente de cometer el error de tipo I, de rechazar la H_0 siendo verdadera.

El segundo factor a tener en cuenta es si el test es de una cola o unilateral, por estar determinada la variable independiente o si es de dos colas o bilateral, por ser admisible tomar como variable independiente cualquiera de las dos variables.

Si se conoce el sentido o la dirección de la relación entre las variables que intervengan en los resultados es preferible el test de una sola cola, ya que faculta, como señala Blalock (1966, pág. 145) para obtener una cola mayor concentrando la región crítica en el extremo apropiado de la región de muestreo.

Otro factor a tener en cuenta es el número de casos de la muestra. Cuando este es muy elevado se puede rechazar fácilmente cualquier H_0 a un nivel de significación muy bajo, por ejemplo, 0,001 y por tanto no se debe dar demasiada importancia a este rechazo.

D y E. CALCULO DEL ESTADISTICO Y DECISION

Esta operación de la prueba de hipótesis estadísticas consiste en efectuar los cálculos necesarios para computar el valor que corresponde, según los datos de la muestra, a la medida estadística que se pretende probar.

Calculado este valor se va a la tabla de valores críticos del test correspondiente y se compara con los que corresponden en ella al nivel de significación elegido y al tipo de prueba de una o de dos colas.

Si en esta comparación el resultado obtenido es menor que el valor crítico de la tabla pertinente, dicho resultado cae fuera de la región crítica y por tanto, no se puede rechazar la H_0.

Si por el contrario es superior, entonces cae dentro de la región crítica y se puede adoptar la decisión de rechazar la hipótesis nula.

CUADRO DE ALTERNATIVAS DE LA DECISION

H_0 / H_0	RECHAZO	ACEPTACION
VERDADERA	Se comete el error de tipo I $p = \alpha$	DECISION CORRECTA
FALSA	DECISION CORRECTA	Se comete el error de tipo II $p = \beta$

DIVISION DE LOS TESTS ESTADISTICOS

Se expone en el siguiente cuadro:

TIPOS DE TESTS ESTADISTICOS

- **Paramétricos.** Exigen requisitos especiales como: variables de intervalo normalidad de las poblaciones e igualdad de su varianza
  - Para contraste con valores dados: De medias, test t; De %, test z
  - Para comparación de dos grupos.: Medias, test t; %, test z
  - Para contraste significatividad coeficientes de correlación: Cada coeficiente tiene su test particular
- **No paramétricos.** No exigen los requisitos de los paramétricos
  - Para tablas de contingencia: Test chi cuadrado; Test de la significación de los cambios
  - Para dos variables ordinales: Prueba de las rachas; Prueba de los signos; Prueba de Smirnov
  - Para dos o más variables ordinales: Prueba de Kruskal y Wallis
- Unilaterales: Si está determinada la variable independiente
- Bilaterales: Si no lo está

CONTRASTE DE MEDIAS Y PROPORCIONES EN RELACION A VALORES DADOS

En este caso la hipótesis que hay que contrastar, la hipótesis nula, consiste en que según una media o proporción obtenida en una muestra, la media o proporción de la población es igual o no difiere significativamente de un valor prefijado que llamaremos a.

Esta hipótesis nula se comprueba mediante la aplicación de las siguientes fórmulas:

CONTRASTE DE RESULTADOS EN RELACION A VALORES DADOS

Fórmulas

Caso de proporciones	Caso de medias
$z = \dfrac{p - a}{\sqrt{a(1-a)/n}}$	$t = \dfrac{a - \bar{x}}{\sigma / \sqrt{n}}$

a = A la media ó la proporción dadas.
p = Al porcentaje obtenido en la muestra que se contrasta.
$\bar{x}$ = A la media obtenida en la muestra.
n = Al tamaño de la muestra.
σ = Desviación típica.

Los valores críticos para contrastar el resultado de la aplicación de la fórmula en el caso de medias, se hallan en la tabla de la distribución de valores de la *t* de Student, tomando n-1 de grados de libertad.

Los valores críticos para contrastar, en el caso de proporciones, la *z* obtenida, se hallan en la tabla de áreas de la curva normal.

El valor de z obtenido representa unidades sigma de dicha tabla de áreas.

Se debe tener en cuenta que un resultado de 1,96 σ es significativa a un nivel del 5 por 100 y otro de 2,58 σ lo es al nivel del 1 por 100.

Respecto al empleo del test *t* o del test *z* se pueden establecer las siguientes normas generales.

En el caso de medias hay que distinguir según se conozca σ, la desviación típica de la población o no. Si se conoce, se emplea el test z, la curva normal. Si no se conoce, en su lugar se trabaja con *s* la desvia-

ción típica de la muestra y se emplea el test *t*, para muestras pequeñas, bien entendido que si el tamaño de la muestra es n > 120 se puede utilizar la curva normal *z* en lugar de *t*.

En el caso de proporciones (Glass, p. 323) se puede considerar la muestra grande y emplear por tanto el test *z*, si n.p o n.(1-p) excede a 5, sea cual sea el menor de estos términos. (Ver ejercicio 213)

COMPARACION DE GRUPOS

Los test de hipótesis también sirven para comparar los resultados obtenidos en la investigación social procedentes de dos muestras diferentes referentes bien a dos poblaciones distintas o bien a la misma población en dos momentos sucesivos.

Con esta comparación se pretende hacer la inferencia o deducción de si los diferentes resultados, obtenidos en las muestras de que se trata, expresan diferencias significativas en las poblaciones de que proceden o, si por el contrario, no es así por ser atribuibles dichas diferencias al azar muestral.

Por ejemplo, supongamos que se han hecho dos encuestas en dos regiones distintas sobre la aceptación del partido x. Supongamos que la muestra de la región A nos da un porcentaje de favorables al partido del 80 por 100 y la de la región B del 60 por 100.

El test de hipótesis nos indica en este caso si estas diferencias de porcentajes en el grado de aceptación del partido X son significativas respecto al total de población de las regiones o no se puede decir basándose en estas muestras que lo sean porque los resultados del test no nos autoriza a rechazar la hipótesis nula.

Es sabido que ésta niega que los datos reflejen una diferencia real y que sean significativos respecto a la población general de las regiones.

En la aplicación de los tests de hipótesis con esta finalidad se pueden distinguir dos casos, según que se quieran comparar resultados expresados en medias o en porcentajes.

Dentro de cada uno de ellos hay a su vez que distinguir, primero, si las muestras son dependientes o independientes, y segundo, si se trata de muestras pequeñas o grandes

Muestras independientes son aquellas en que no hay asociación ninguna en las poblaciones de que proceden respecto a la característica de que se trate.

Muestras dependientes son aquellas en que existe tal asociación, como por ejemplo en el caso de que se trate de dos muestras extraídas de poblaciones formadas por pares de gemelos o de varones y mujeres procedentes de matrimonios.

Blalock (1966, pág. 160) aunque señala que hay que ser cauto en el empleo de las reglas prácticas, estima que es grande siempre o casi siempre una muestra de N > 100, y también de n > 50 si se tiene evidencia empírica de que la desviación de la población en relación a la normalidad no es importante. Por el contrario, una muestra de N < 30 debe estimarse siempre como pequeña, salvo que se esté seguro de la normalidad de la población.

Las fórmulas a aplicar en cada caso son las que se expresan en el cuadro que sigue.

En ellas como en el caso anterior, los resultados *z* y *t*, vienen expresados en unidades sigma del área de la curva normal o en unidades de la distribución *t* de Student y $n_1 - 1 + n_2 - 1$ grados de libertad.

Si el *z* o *t* obtenido es mayor que el *z* o *t* crítico de la tabla correspondiente se rechaza la H_0 y si es menor no, o sea, se acepta ésta.

En general, en las fórmulas se aplica el método de la razón crítica que consiste en dividir las diferencias de medias o porcentajes por el error muestral, de dichas diferencias.

Fórmulas

Tipos de muestras	**Para MEDIAS**	**Para PROPORCIONES**
Grandes independientes	$t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$	$z = \dfrac{p_1 - p_2}{\sqrt{\dfrac{p_1 \cdot q_1}{n_1} + \dfrac{p_2 \cdot q_2}{n_2}}}$
Pequeñas independientes	$t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} \cdot \dfrac{n_1 + n_2}{n_1 \cdot n_2}}}$	$z = \dfrac{p_1 - p_2}{\sqrt{\left(\dfrac{p_1 + p_2}{n_1 + n_2}\right)\left(1 - \dfrac{p_1 + p_2}{n_1 + n_2}\right)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$

Dependientes	Es la misma que en las independientes, con la diferencia de que se multiplica el radicando por el factor $(1 - r^2_{xy})$	$z = \frac{d - a}{\sqrt{d + a}}$	a b / c d

n_1 y n_2 = Son los tamaños de las muestras comparadas.
$\bar{x}_1$ y $\bar{x}_2$ = Son las medias que se comparan.
p_1 y p_2 = Son los porcentajes que se comparan.
σ_1^2 y σ_2^2 = Son las varianzas de las poblaciones que se comparan.
s_1 y s_2 = Son las varianzas de las muestras que se comparan.
q_1 y q_2 = Son $1 - p_1$ y $1 - p_2$ respectivamente.
$1\ r^2_{xy}$ = Es el coeficiente de correlación entre las variables dependiente e independiente.

De los casos expuestos presenta especial interés, el referente a la comparación de las medias de dos grupos dependientes conectados par a par.

Según señala Mueller (1970, pág. 416), se encuentran por lo menos tres tipos de estos grupos: 1. Grupos apareados a efectos de un diseño experimental. 2. Grupo de pares naturales como el de los maridos por una parte y las mujeres por otra. 3. Observaciones del mismo grupo en diferentes tiempos.

Además este procedimiento es el que se emplea a menudo en la construcción de las escalas tipo Likert para contrastar las puntuaciones en los mismos ítems, del 25 % del grupo que obtenga las puntuaciones más elevadas y el 25 % del grupo que las obtenga más bajas.

En este caso, ya que hay frecuencia en los items, la fórmula a aplicar es la siguiente:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{n\,(n-1)}}}$$

En el caso general, la fórmula es:

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \frac{\sigma\, d}{\sqrt{N}} \qquad t = \frac{\bar{x}_1 - \bar{x}_2}{\sigma_{\bar{x}_1 - \bar{x}_2}}$$

En la que σ d es la desviación típica de la distribución de las diferencias entre los individuos apareados y N el número de apareamientos.

(Ver ejercicios 205 y 211 a 216)

CONTRASTE DE LOS COEFICIENTES DE CORRELACION

Para hallar si los coeficientes de asociación obtenidos de muestras son significativamente distintos de cero respecto a la población, el procedimiento a seguir es el acabado de exponer.

Por tanto, esta cuestión no presenta dificultad, sobre todo si se conoce la distribución de muestreo y existen tablas que den en las mismas unidades de los coeficientes los valores críticos de H_0, como ocurre en el caso del coeficiente r de Pearson y la gamma de Goodmann y Kruskal.

Una exposición de este punto, y en general bastante completa respecto a los distintos test de hipótesis, se puede encontrar en Glass (1973, págs. 307 y ss.) y en Ackoff (1973, págs. 185 y ss.). (Ver ejercicios 208, 209 y 206).

Respecto al coeficiente phi existe un contraste muy fácil, pues se reduce a la aplicación de la siguiente fórmula:

$$z = \sqrt{n \quad \varphi}$$

en la que n, es el número de elementos de la muestra, y φ el coeficiente obtenido. (Ver ejercicio 207)

Nos da el resultado en valores de la distribución *z* citada, con la que hay que contrastar el resultado que se obtenga.

También se pueden contrastar las varianzas con valores dados o comparándolas entre sí. (Vid. Glass, 1973, págs. 300 y ss.)

ANALISIS DE LA VARIANZA

Los métodos expuestos anteriormente igual que el método t se refieren a la comparación de dos grupos distintos.

Pero en la investigación social es muy frecuente el caso de tener que comparar entre sí varios grupos. Este problema se podría solucionar comparando todos los grupos de dos en dos, pero esto, además de ser muy penoso, multiplicaría los errores que siempre existen.

El análisis de varianza soluciona este problema, pues permite la comparación conjunta de más de dos a la vez.

Este análisis se basa en comparar estadísticamente la variabilidad o las diferencias que se dan entre los grupos con las diferencias que existen dentro de cada grupo. Si las diferencias son significativas hay que concluir que se trata de poblaciones distintas en relación a la característica investigada.

De aquí que F, el resultado del análisis de varianza, sea igual al siguiente cociente:

$$\frac{\text{Estimación de la varianza basada en la diferencia de las medias de los grupos}}{\text{Estimación de la varianza basada en la varianza entre las muestras o entre los diferentes grupos.}}$$

El primer factor de esta división no lo explica el análisis de varianza, por eso se llama varianza inexplicada. El segundo sí y recibe el nombre de varianza explicada.

El análisis de varianza se emplea también para comprobar la asociación entre una o varias escalas nominales y otra de intervalo y sobre todo en los experimentos.

En éstos, cuando se han realizado múltiples ensayos del mismo experimento, es una técnica muy valiosa para contrastar la significatividad de las variaciones que se observen entre los distintos ensayos.

El objetivo del análisis de varianza es comprobar la significatividad de las diferencias entre las medias de diversas muestras. Por ello, se puede considerar como una extensión del test de las diferencias de las medias.

La hipótesis nula que se contrasta en el análisis de varianza es que no hay diferencia entre los distintos grupos analizados y que las variaciones observadas son debidas al azar.

Si esta hipótesis nula es correcta, la variación observada dentro de los grupos no deberá ser muy superior a la observada entre los grupos y deben caer dentro de los límites de variación fijados en la distribución muestral F.

Se pueden distinguir diversos tipos de análisis de varianza. Entre ellos el más común y al que me limito aquí es el análisis de varianza simple en el que se tiene, además de la variable de intervalo, una variable nominal. También es corriente el análisis de varianza bimodal, en el que las variables nominales son dos en lugar de una sola.

Los supuestos en que se funda son: 1. La normalidad de la curva de distribución. 2. La independencia de las muestras aleatorias. 3. La igualdad de las desviaciones típicas. 4. El carácter sumativo de los efectos de la variación total y no multiplicativo.

Hay que tener en cuenta que el análisis de varianza no dice si las diferencias de los grupos dos a dos son o no significativas. Para saberlo, si es necesario, habría que emplear el método t u otro aplicable.

OPERACIONES QUE COMPRENDE

En el análisis de varianza, una vez se ha formulado la hipótesis nula, para contrastarla se procede de la siguiente forma:

1. Se establece la tabla de los grupos de modo que éstos figuren por columnas.

2. Se hallarán las sumas de cada columna y la suma total de todas ellas.

3. Se hallan los cuadrados de cada uno de los datos de todas las columnas y se suman por columnas.

4. Se obtiene el término de corrección dividiendo el cuadrado del total general de cada tabla por N = al número de datos de la tabla.

5. Se halla la suma de cuadrados total, que es la diferencia entre la suma de los cuadrados de los datos de todas las columnas (3) y el factor de correción.

6. Se halla la suma de cuadrados intergrupo, para lo cual se elevan al cuadrado cada una de las sumas de las columnas, se dividen los cuadrados por el número de datos o elementos que forman cada columna y de la suma de los cocientes se resta el factor de corrección.

7. Se halla la suma de cuadrados intragrupo, que es igual a la diferencia entre la suma de cuadrados total (5), menos la suma de cuadrados intergrupo (6).

8. Hallados estos datos, se ordenan los resultados en la tabla siguiente, en la que las medias cuadráticas es el resultado de dividir las sumas de cuadrados intergrupo e intragrupo por los grados de libertad, referentes al total de grupos y a la tabla en general, respectivamente, y el valor de F es el cociente de las medias cuadráticas:

Fuente de variación	Suma de cuadrados	Grados de libertad	Medias cuadrática	Valor de F
Intergrupo				
Intragrupo				

9. Hallado F., se compara su valor con el que den las tablas para los grados de libertad intergrupo e intragrupo en cuestión y se procede a

aceptar o rechazar la hipótesis nula de manera similar a los casos anteriores (Ver ejercicio 205).

TEST χ^2

Este tipo de test es muy importante y muy usado en los estudios sociológicos. Se utiliza especialmente con relación a variables cualitativas de escala ordinal o nominal, o sea, en aquellas en que sus dimensiones o categorías sólo indican distinción u orden, pero no intervalos iguales.

Este test es uno de los instrumentos principales para el análisis estadístico de las tablas de contingencia. En la década del 70 este análisis ha experimentado un gran desarrollo (Swafford 1980, 664) con la aplicación del cálculo logarítmico, que se realiza en los tipos de análisis llamados log-linear y logit. Sin embargo, estas técnicas aunque emplean los logarítmos para hallar las probabilidades, siguen basándose en la utilización de este test. (Sobre estas técnicas se pueden consultar el artículo citado y Haberman, 1978).

Consiste en la comparación entre las frecuencias realmente obtenidas, que son los valores de cada dimensión de la variable, hallados en la investigación y reflejados en las tablas, y las frecuencias teóricas o normales, que serían las que corresponderían a cada categoría o casilla de la tabla si su valor siguiese cualquier norma teórica adoptada, por ejemplo, curva normal, igualdad absoluta o igual distribución proporcional de las frecuencias de cada variable en todos los casos.

Supuesto esto, el χ^2 es igual a la suma de los cuadrados de las diferencias entre las frecuencias realmente obtenidas y las frecuencias normales, divididas por su correspondiente frecuencia normal.

Por tanto, la fórmula es la siguiente:

$$\chi^2 = \Sigma \frac{(f_o - f_n)^2}{f_n}$$

Fórmula abreviada:

$$\chi^2 = \Sigma \frac{f_o^2}{f_n} - N$$

En la que f_o, son las frecuencias realmente obtenidas y f_n, las frecuencias teóricas normales.

Para hallar las frecuencias teóricas normales en el caso de que se adopte la igual distribución proporcional, se multiplica el total de la fila por el total de la columna, que corresponda a cada valor o dato, y se divide por N, o el total de individuos de la muestra.

Los valores de las tablas para contrastar los valores obtenidos de χ^2 están dados en función de los llamados grados de libertad. Por tanto, es necesario conocer éstos para aplicar la tabla.

Si el número total de individuos de la muestra de una columna de la tabla, es por ejemplo 100, y hay seis casillas, la condición inexcusable de los valores de éstas casillas es que sumen 100, lo que supone que cualquiera de las seis casillas menos una, podrá tener otro número cualquiera. Por tanto, se dice que tenemos 6-1 cinco grados de libertad.

Su fórmula, pues, es N - 1 en las filas y M - 1 en las columnas, y para la tabla completa (N - 1) (M - 1).

Este test no puede usarse cuando alguna o algunas de las frecuencias esperadas son inferiores a cinco.

En este caso conviene agrupar categorías similares para que esto no suceda.

Asimismo, para el caso de tablas de contingencia de 2 x 2 o de un grado de libertad se puede aplicar la fórmula especial siguiente:

$$\chi^2 = \frac{N\left(|a \cdot d - b \cdot c| - \frac{N}{2}\right)^2}{(a+b)\ (c+d)\ (a+c)\ (b+d)}$$

a	b
c	d

Esta fórmula comprende la corrección de Yates, que se debe hacer cuando se aplica la fórmula general, en este caso de un grado de libertad.

Consiste en aumentar en 0,5 las frecuencias empíricas que son menores que sus frecuencias teóricas y en disminuir 0,5 las frecuencias empíricas que son mayores que sus correspondientes teóricas.

El test chi cuadrado se puede utilizar como test de significación para contrastar la hipótesis nula de que las diferencias observadas en la muestra no se pueden extender al universo o población, o bien como test de

independencia para comprobar la asociación entre las variables de la tabla, o sea, para ver si se puede rechazar o no la hipótesis nula de la independencia entre ambas variables.

Hay que tener en cuenta que, en este caso, chi cuadrado sólo indica la existencia de asociación, pero no mide el grado de ésta.

Una medida fundada en χ^2 del grado de relación, asociación o dependencia de las clasificaciones en una tabla de contingencia es dada por:

$$c = \sqrt{\frac{\chi^2}{\chi^2 + N}}$$

que se llama coeficiente de contingencia. A mayor valor de c, mayor grado de asociación. El valor máximo de K viene dado por:

$$\sqrt{k - 1/k}$$

k = Número de filas y columnas.

El coeficiente de contingencia tiene el inconveniente, como se deriva de lo dicho, de no variar de 0 a 1. Este inconveniente se puede evitar aplicando en su lugar el coeficiente V de Cramer o el T de Tschprow.

Su fórmula es la que sigue:

$$V = \sqrt{\frac{\chi^2}{N.\ \text{min.}\ (h-1;\, c-1)}}$$

$$T^2 = \frac{C^2}{(1 - C^2)\sqrt{(h-1)(c-1)}}$$

Min. $(h - 1; c - 1)$ es el menor valor de ellos, siendo h. el número de hilera o de filas y c. el número de columnas. C^2 es el coeficiente de contingencia.

Por último es de recordar que chi cuadrado también se utiliza, como vimos en el estudio de las escalas, para comprobar si una determinada distribución empírica difiere o no significativamente de la distribución que representa la curva normal. Se trata de lo que los ingleses llaman comprobación de la "goodness of fit", bondad del ajuste.

El test χ^2 prueba la H_0 en relación con la tabla en su conjunto pero no respecto a los diversos subconjuntos que se pueden formar dentro

de la tabla general. Para contrastar la H_0 en relación a éstos Maxwell (1961, 55) expone un procedimiento, así como para comparar las diferencias entre diferentes investigaciones y ver si son significativas. (Ver ejercicios 204 y 205)

EJERCICIOS

204

Se ha estudiado la asociación entre el sexo y la filiación política mediante una pequeña muestra, N = 30, de 17 hombres y 13 mujeres.

Los resultados se expresan en la siguiente tabla:

	FILIACION POLITICA		
Sexo	*No afiliados a ningún partido*	*Afiliados a algún partido*	*Total*
Hombres	7	10	17
Mujeres	10	3	13
Total	17	13	30

1. Formular la hipótesis nula.

2. Contrastarla mediante la prueba del χ^2, empleando para ello la siguiente fórmula especial para tablas de contingencias de 2 x 2:

$$\chi^2 = \frac{N\left[(a \cdot d - b \cdot c) - \frac{N}{2}\right]^2}{(a+b)\cdot(c+d)\cdot(a+c)\cdot(b+d)}$$

en la que a y b son los valores de la primera fila en este caso, 7 y 10, y c y d los de la segunda, 10 y 3.

N es el número total de individuos de la muestra.

RESPUESTA

1. Hipótesis nula: No existe en la población asociación entre el sexo y la afiliación política. La asociación muestral es debida al azar.

2. Prueba de Ji cuadrado. Aplicando la fórmula tenemos:

$$\frac{30\left[(7\times 3-10\times 10)-\frac{30}{2}\right]^2}{(7+10)\times(10+3)\times(7+10)\times(10+3)}=\frac{122.880}{48.441}=2{,}52$$

El valor de ji cuadrado en la tabla para un grado de libertad $(2-1)\times(2-1)=1$ y para un nivel de significación de 0,05 es de 3,84, es decir, superior al obtenido. Ello indica que no se puede rechazar la hipótesis nula ni deducir de la muestra obtenida que exista una asociación significativa entre el sexo y la afiliación política en la población de que procede la muestra.

205

Dadas las siguientes tablas:

A).

PROVINCIAS

Industrializadas		*De agricultura autonónoma*		*De agricultura asalariada*	
Barcelona	31	Avila	21	Cadiz	16
Guipuzcoa	12	Burgos	32	Córdoba	36
Vizcaya	29	Palencia	31	Jaén	19
Oviedo	28	Cuenca	21	Huelva	21
Santander	15	Segovia	20	Málaga	21
		C. Real	29		

Las cifras son los valores de un índice de movilidad vertical (ascendente), año de 1966, construido con datos de encuesta, tomando como base el cambio ocupacional intergeneracional.

Representan el porcentaje sobre el total de la población activa encuestada que declaró en la encuesta tener una ocupación superior a la de sus padres.

(Fuente: J. Díez Nicolás. Segundo estudio. Tres estudios para un sistema de indicadores sociales. Foessa, 1967, ps 227. y ss.)

A)

VARIACION DE x^2

Niveles de significación

Grados de libertad	P = .99	.98	.95	.90	.80	.70	.50	.30	.20	.10	.05	.02	.01	.001
1	.000157	.000628	.00393	.0158	.0642	.148	.455	1.074	1.642	2.706	3.841	5.412	6.635	10.827
2	.0201	.0404	.103	.211	.446	.713	1.386	2.408	3.219	4.605	5.991	7.824	9.210	13.815
3	.115	.185	.352	.584	1.005	1.424	2.366	3.665	4.642	6.251	7.815	9.837	11.341	16.268
4	.297	.429	.711	1.064	1.649	2.195	3.357	4.878	5.989	7.779	9.488	11.668	13.277	18.465
5	.554	.752	1.145	1.610	2.343	3.000	4.351	6.064	7.289	9.236	11.070	13.388	15.086	20.517
6	.872	1.134	1.635	2.204	3.070	3.828	5.348	7.231	8.558	10.664	12.592	15.033	16.812	22.457
7	1.239	1.564	2.167	2.833	3.822	4.671	6.346	8.383	9.803	12.017	14.067	16.622	18.475	24.322
8	1.645	2.032	2.733	3.490	4.594	5.527	7.344	9.524	11.030	13.362	15.507	18.168	20.090	26.125
9	2.088	2.532	3.325	4.168	5.380	6.393	8.343	10.656	12.242	14.684	16.919	19.679	21.666	27.877
10	2.558	3.059	3.940	4.865	6.179	7.267	9.342	11.781	13.442	15.987	18.307	21.161	23.209	29.588
11	3.053	3.609	4.575	5.578	6.989	8.148	10.341	12.899	14.631	17.275	19.675	22.618	24.725	31.264
12	3.571	4.178	5.226	6.304	7.807	9.034	11.340	14.011	15.812	18.549	21.026	24.054	26.217	32.909
13	4.107	4.765	5.892	7.042	8.634	9.926	12.340	15.119	16.985	19.812	22.362	25.472	27.688	34.528
14	4.660	5.368	6.571	7.790	9.467	10.821	13.339	16.222	18.151	21.064	23.685	26.873	29.141	36.123
15	5.229	5.985	7.261	8.547	10.307	11.721	14.339	17.322	19.311	22.307	24.996	28.259	30.578	37.697
16	5.812	6.614	7.962	9.312	11.152	12.624	15.338	18.418	20.465	23.542	26.296	29.633	32.000	39.252
17	6.408	7.255	8.672	10.085	12.002	13.531	16.338	19.511	21.615	24.769	27.587	30.995	33.409	40.790
18	7.015	7.906	9.390	10.865	12.857	14.440	17.338	20.601	22.760	25.989	28.869	32.346	34.805	42.312
19	7.633	8.567	10.117	11.651	13.716	15.352	18.338	21.689	23.900	27.204	30.144	33.687	36.191	43.820
20	8.260	9.237	10.851	12.443	14.578	16.266	19.337	22.775	25.038	28.412	31.410	35.020	37.566	45.315
21	8.897	9.915	11.591	13.240	15.445	17.182	20.337	23.858	26.171	29.615	32.671	36.343	38.932	46.797
22	9.542	10.600	12.338	14.041	16.134	18.101	21.337	24.939	27.301	30.813	33.924	37.659	40.289	48.268
23	10.196	11.293	13.091	14.848	17.187	19.021	22.337	26.018	28.429	32.007	35.172	38.968	41.638	49.728
24	10.856	11.992	13.848	15.659	18.062	19.943	23.337	27.096	29.553	33.196	36.415	40.270	42.980	51.179
25	11.524	12.697	14.611	16.473	18.940	20.867	24.337	28.172	30.675	34.382	37.652	41.566	44.314	52.620
26	12.198	13.409	15.379	17.292	19.820	21.792	25.336	29.246	31.795	35.563	38.885	42.856	45.642	54.052
27	12.879	14.125	16.151	18.114	20.703	22.719	26.336	30.319	32.912	36.741	40.113	44.140	46.963	55.476
28	13.565	14.847	16.928	18.939	21.588	23.647	27.336	31.391	34.027	37.916	41.337	45.419	48.278	56.893
29	14.256	15.574	17.708	19.768	22.475	24.577	28.336	32.461	35.139	39.087	42.557	46.693	49.588	58.302
30	14.953	16.306	18.493	20.599	23.364	25.508	29.336	33.530	36.250	40.256	43.773	47.962	50.892	59.703

Si los grados de libertad son más de 30 se utiliza como contraste la tabla de arcas de la curva normal, seguir la siguiente fórmula: $z=\sqrt{2\chi^2}\cdot\sqrt{2n-1}$

B). En una investigación realizada se han obtenido los siguientes resultados respecto al espectáculo preferido por las profesionales que se indican:

Espectáculo preferido	*Propietarios agrícolas*	*Empresarios*	*Profes. liberales*	*Obreros y empleados*	*Total*
Fútbol	70	100	90	170	430
Toros	200	90	160	190	640
Cine	360	200	250	30	840
Teatro	30	10	40	10	90
Totales	660	400	540	400	2.000

C). Se ha aplicado para su prueba una escala tipo Likert a un grupo de 200 individuos.

Los resultados en relación al item 5, para el 25 % de los que alcanzaron la puntuación máxima en la escala y el 25 % la mínima, subgrupos superior e inferior, fueron los siguientes:

Puntuación dada al ítem		*Grupo inferior*	*Grupo superior*
4		2	20
3		4	6
2		6	4
1		20	4
0		8	6
	N =	40	40

Se pide:

1) Indicar las variables que forman las tablas y su tipo según su escala.

RESPUESTA

Tabla A. Hay dos variables. La primera, el tipo de economía provincial predominante, de tipo independiente, y nominal con tres categorías.

La segunda variable está formada por los valores del índice de movilidad intergeneracional correspondiente a cada provincia. Es una variable dependiente y de escala de intervalo.

Tabla B. Variables: Ocupación y tipo de espectáculo preferido, las dos de escala nominal.

Tabla C. Variables: Puntuación en la escala tipo Likert, de escala ordinal, aunque se puede considerar como de intervalo a efectos de aplicaciones de ciertas pruebas.

La segunda variable es la formada por los dos subgrupos con puntuaciones máximas y mínimas, de escala nominal.

2) Expresar las hipótesis de investigación que suponen las tablas.

RESPUESTA

Tabla A. Existe diferencia en las variaciones provinciales del índice de movilidad social, según los distintos tipos de actividad económica provincial.

Tabla B. El tipo de espectáculo preferido es distinto según las diferentes profesiones.

Tabla C. Las puntuaciones en la escala de los subgrupos con puntuaciones máximas y mínimas son significativamente diferentes.

3) Señalar el test estadístico para contrastar H_0 aplicable en cada tabla.

RESPUESTA

A) El análisis de varianza, por ser una variable nominal e independiente, y otra de razón y por comprender la variable nominal más de dos categorías o grupos.

B) El test chi cuadrado, por ser la tabla de contingencia y las variables nominales.

C) El test t, por ser sólo dos grupos y porque se puede considerar a efectos de esta prueba la variable dependiente de intervalo.

4) Formular la H_0 y contrastarla al nivel de significación del 5 % calculando el test adecuado.

RESPUESTA

Tabla A. H_0. No existen diferencias en las variaciones provinciales del índice de movilidad social, según los distintos tipos de actividad económica.

TABLA A

Fórmulas:

$$SCT = \Sigma X_{ij}^2 - T_G^2/N$$

$$SCE = \Sigma T_i^2/n_i - T_G^2/N;$$

$$SCD = SCT - SCE.$$

Tabla de datos

	Industriales		*Agri. autó.*		*Agri. asala.*		*Totales*
	x_1	x_1^2	x_2	x_2^2	x_3	x_3^2	
	31	961	21	441	16	256	
	12	144	32	1024	36	1296	
	29	841	31	961	19	361	
	28	784	21	441	21	441	
	15	225	20	400	21	441	
			29	841			
n_i	5		6		5		N = 16
T_i	115		154		113		T_G = 382
T_i^2	13225		23716		12769		
T_i^2/n_i	2645		3952.66		2553.8		9151.46
Σx_1^2	2955		4108		2795		9858

Fuente de variación	*suma de cuadrados*	*Grados de libe.*	*Varianzas*	*F*
Entre grupos	SCE = 9151.46 – $(382)^2/16$ = = 31.21	2	15.605	.28
Dentro de grupos o residual	*SCD* = 737.75 – 31.21 = = 706.54	13	54.34	
Total	SCT = 9858 – $(382)^2/16$ = = 737.75	15	49.18	

El F observado tan pequeño, menor que la unidad, indica de antemano la nula significatividad del análisis de varianza realizado, como se confirma hallando el F crítico, $F_{(13,\,2,\,.05)} = 19.42$.

Se acepta, pues, la hipóstesis nula. Las variables, tipo de economía provincial e índice de movilidad social intergeneracional, son independientes. No hay diferencias en dicha movilidad entre los tipos de provincias distinguidas.

TABLA B

H_0. El tipo de espectáculo preferido no es distinto según las diferentes profesiones.

Cálculo:

Propietarios agrícolas:

$$\frac{660 \times 430}{2.000} = \frac{283.800}{2.000} = 141{,}9; \quad \frac{660 \times 640}{2.000} = \frac{422.400}{2.000} = 211{,}2$$

$$\frac{660 \times 840}{2.000} = \frac{554.400}{2.000} = 277{,}20; \quad \frac{660 \times 90}{2.000} = \frac{59.400}{2.000} = 29{,}7$$

Empresarios y obreros (al ser la suma de las columnas iguales, los valores teóricos son los mismos para estas dos categorías):

$$\frac{400 \times 430}{2.000} = \frac{172.000}{2.000} = 86; \quad \frac{400 \times 640}{2.000} = \frac{256.000}{2.000} = 128$$

$$\frac{400 \times 840}{2.000} = \frac{336.000}{2.000} = 168; \quad \frac{400 \times 90}{2.000} = \frac{36.000}{2.000} = 18$$

Profesiones liberales:

$$\frac{540 \times 430}{2.000} = \frac{232.200}{2.000} = 116{,}1; \quad \frac{540 \times 640}{2.000} = \frac{345.600}{2.000} = 172{,}80$$

$$\frac{540 \times 840}{2.000} = \frac{453.600}{2.000} = 226{,}8; \quad \frac{540 \times 90}{2.000} = \frac{48.600}{2.000} = 24{,}30$$

e_n	t_n	$e_n - t_n$	$(e_n - t_n)^2$	$\frac{(e_n - t_n)^2}{t_n}$
70	142	— 72	5.184	36,50
200	211,2	— 11,2	125,44	0,59
360	277,2	82,8	6.855,84	24,73
30	29,7	0,3	0,09	0,003
100	86	14	196	2,27
90	128	— 38	1.444	11,28
200	168	32	1.024	6,09
10	18	— 8	64	3,55

90	116,1	— 26,1	681,21	5,86
160	172,8	— 12,8	169,84	0,94
250	226,8	23,2	538,24	2,37
40	24,3	15,7	246,49	10,14
170	86	84	7.056	82,04
190	128	62	3.844	30,03
30	168	— 138	19.044	113,35
10	18	— 8	64	33,55
				333,29

Grados de libertad (4 columnas – 1) x (4 filas – 1) = 9.

El valor de ji cuadrado en la tabla para 9 grados de libertad al nivel de significación del 1 por 1.000 es de 27,877. El resultado obtenido es muy superior, luego es preciso rechazar la hipótesis nula, al nivel máximo de significación del 1 por 1.000. De los resultados obtenidos se puede inferir que existe una asociación significativa entre el espectáculo preferido y la ocupación profesional en la población considerada. Sin embargo, el resultado obtenido muestra un ejemplo de lo fácil que es, como se indicó, rechazar la H_0 cuando el número de casos es muy elevado.

TABLA C

1. H_0 = No hay diferencia entre los grupos superior e inferior. La obtenida se explica por el azar o la fluctuación muestral.

2. El test apropiado es el t, según la fórmula siguiente:

$$t = \frac{\bar{x}_2 - \bar{x}_1}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{n\,(n-1)}}}$$

Sustituyendo la fórmula se tiene:

Puntuación dada al grupo inferior, x_1	f_1	$f_1 \cdot x_1$	$f_1 \cdot x_1^2$
4	2	8	32
3	4	12	36
2	6	12	24
1	20	20	20
0	8	8	0
	40	52	112

Grupo superior:

Puntuación dada x_2	*Frec. f_2*	$f_2 \cdot x_2$	$f_2 \cdot x_2^2$
4	20	80	320
3	6	18	54
2	4	8	16
1	4	4	4
0	6	6	0
	40	110	394

$$\bar{x}_1 = \frac{52}{40} = 1{,}3; \quad \bar{x}_2 = \frac{110}{40} = 2{,}75; \quad \sigma_1^2 = 112 - \frac{(52)^2}{40} = 44{,}4$$

$$\sigma_2^2 = 394 - \frac{(110)^2}{40} = 91{,}5; \quad t = \frac{2{,}75 - 1{,}13}{\sqrt{\frac{91{,}5 + 44{,}4}{40 \cdot 39}}} = 4{,}41$$

3. En las tablas t al nivel de 0,05 para 78 grados de libertad vale 2. El t obtenido es mayor, 4,91, luego se rechaza la hipótesis nula.

La respuesta de los subgrupos inferior y superior al item 5 de la escala es estadísticamente significativamente diferente.

(Basado en Mayntz, 1975, pág. 79).

206

Dado el coeficiente de asociación Rho de Spearman obtenido en el ejercicio 191 de 0,21, se pide decidir si es o no significativo, N = 10, al nivel de significación del 0,5.

RESPUESTA

H_0 = El coeficiente obtenido no es significativo. Se explica por la variación muestral.

El test apropiado es z, según la siguiente fórmula:

$$z = \frac{\rho - 0}{1/\sqrt{N-1}} \qquad z = \frac{0{,}21}{1/\sqrt{5}} = \frac{0{,}21}{1/2{,}23} = \frac{0{,}21}{0{,}45} = 0{,}46$$

El valor crítico de Z al 0,05 es 1,96. El Z obtenido es menor: 0,46 Por tanto no se puede rechazar la H_0.

La variación observada no es estadísticamente significativa. Se explica por el error muestral.

207

Dado un coeficiente φ obtenido de 0,74, se pide decidir si es o no significativo, N = 170, al nivel de significación de 0,01.

RESPUESTA

H_0 = El coeficiente obtenido no es significativo para la población de donde proceda la muestra. Se explica por la variación muestral;

El test apropiado es Z según la siguiente fórmula:

$$z = \sqrt{n} \cdot \varphi; \quad z = \sqrt{170} \cdot 0{,}75 = 9{,}77\ \sigma$$

El Z obtenido es muy superior al crítico. Luego se rechaza la H_0 que postulaba no existía asociación en la población de donde procede la muestra.

VALORES SIGNIFICATIVOS DE G PARA CONTRASTAR H_0

	Test bilateral		*Test unilateral*	
N	$\alpha = 0{,}05$	$\alpha = 0{,}01$	$\alpha = 0{,}05$	$\alpha = 0{,}01$
4			1,000	
5	1,000		0,800	1,000
6	0,867	1,000	0,733	0,867
7	0,714	0,905	0,619	0,810
8	0,643	0,786	0,571	0,714
9	0,556	0,722	0,500	0,667
10	0,511	0,644	0,497	0,600
11	0,491	0,600	0,418	0,564
12	0,455	0,576	0,394	0,545
13	0,436	0,564	0,359	0,513
14	0,407	0,516	0,363	0,473
15	0,390	0,505	0,333	0,467
16	0,383	0,483	0,317	0,433
17	0,368	0,471	0,309	0,426

18	0,346	0,451	0,294	0,412
19	0,333	0,439	0,287	0,392
20	0,326	0,421	0,274	0,379
21	0,314	0,410	0,267	0,371
22	0,307	0,394	0,264	0,359
23	0,296	0,391	0,257	0,352
24	0,290	0,377	0,246	0,341
25	0,287	0,367	0,240	0,333
26	0,280	0,360	0,237	0,329
27	0,271	0,356	0,231	0,322
28	0,265	0,344	0,228	0,312
29	0,261	0,340	0,222	0,310
30	0,255	0,333	0,218	0,301
31	0,252	0,325	0,213	0,295
32	0,246	0,323	0,210	0,290
33	0,242	0,314	0,205	0,288
34	0,237	0,312	0,201	0,280
35	0,234	0,304	0,197	0,277
36	0,232	0,302	0,194	0,273
37	0,228	0,297	0,192	0,267
38	0,223	0,292	0,189	0,263
39	0,220	0,287	0,188	0,260
40	0,218	0,285	0,185	0,256
	1,960	2,576	1,645	2,326

208

Dados el resultado G obtenidos en el ejercicio 194 sobre:

La asociación entre los rangos de obreros y de profesionales respecto a las cualidades deseables para sus hijos.

Se pide formular la hipótesis nula, contrastarla utilizando la tabla G y, a un nivel de significación del 0,5 por 100, interpretar los resultados.

RESPUESTA

Hipótesis nula: Los resultados obtenidos son debidos al azar y no existe asociación positiva entre las ordenaciones de las cualidades que se estiman importantes para los hijos por las familias de trabajadores y de profesionales.

Contraste de la hipótesis: La relación entre las dos ordenaciones se puede considerar bilateral.

Valor de G obtenido en el ejercicio: 0,09.

Valor de G según la tabla para N = 12 rangos y a un nivel de significación de 0,05 y test bilateral 0,455.

Interpretación: Este valor en la tabla es mucho mayor que el de G obtenido empíricamente, luego no se puede rechazar la hipótesis nula y hay que concluir que no se puede inferir de la muestra que exista en la población considerada una asociación positiva entre ambas ordenaciones.

VALORES SIGNIFICATIVOS DE R PARA CONTRASTAR H_o

	TEST BILATERAL		TEST UNILATERAL	
gl	$\alpha = 0{,}05$	$\alpha = 0{,}01$	$\alpha = 0{,}05$	$\alpha = 0{,}01$
1	0,997	0,9999	0,988	0,9995
2	0,950	0,990	0,900	0,980
3	0,878	0,959	0,805	0,934
4	0,811	0,917	0,729	0,882
5	0,754	0,874	0,669	0,833
6	0,707	0,834	0,622	0,789
7	0,666	0,798	0,582	0,750
8	0,632	0,765	0,549	0,716
9	0,602	0,735	0,521	0,685
10	0,576	0,708	0,497	0,658
11	0,553	0,684	0,476	0,634
12	0,532	0,661	0,458	0,612
13	0,514	0,641	0,441	0,592
14	0,497	0,623	0,426	0,574
15	0,482	0,606	0,412	0,558
16	0,468	0,590	0,400	0,542
17	0,456	0,575	0,389	0,528
18	0,444	0,561	0,378	0,516
19	0,433	0,549	0,369	0,503
20	0,423	0,537	0,360	0,492
21	0,413	0,526	0,352	0,482
22	0,404	0,515	0,344	0,472
23	0,396	0,505	0,337	0,462
24	0,388	0,496	0,330	0,453
25	0,381	0,487	0,323	0,445
26	0,374	0,479	0,317	0,437
27	0,367	0,471	0,311	0,430
28	0,361	0,463	0,306	0,423
29	0,355	0,456	0,301	0,416
30	0,349	0,449	0,296	0,409
35	0,325	0,418	0,275	0,381
40	0,304	0,393	0,257	0,358
45	0,288	0,372	0,243	0,338
50	0,273	0,354	0,231	0,322
60	0,250	0,325	0,211	0,295
70	0,232	0,303	0,195	0,274
80	0,217	0,283	0,183	0,256
90	0,205	0,267	0,173	0,242
100	0,195	0,254	0,164	0,230

209

En una muestra de 82 individuos la correlación r de Pearson entre los coeficientes inteligencia y agresividad fue de 0,59.

Se pide:

1. *Formular la hipótesis nula.*
2. *Indicar si se trata de un test unilateral o bilateral.*
3. *Contrastar la hipótesis nula.*
4. *Interpretar el resultado.*

RESPUESTA

1. Hipótesis nula: No existe una asociación positiva entre los coeficientes de inteligencia y agresividad.

2. Tipo de test a aplicar: Se puede postular que la agresividad influye en la inteligencia y al revés, por ello se debe aplicar el test bilateral.

3. Contraste de la hipótesis:

Grados de libertad $(N - 2) = 82 - 2 = 80$.

La tabla da un valor de 0,283 para el test bilateral y el nivel de significación 0,01.

Interpretación: Como el coeficiente r de Pearson obtenido de la muestra es muy superior, 0,59, se rechaza la hipótesis nula y se postula que la muestra proviene de una población en la que existe asociación entre la agresividad y el coeficiente de inteligencia, a un nivel de significación del 0,01, es decir, con una probabilidad de 99 por 100.

210

Dados los resultados del ejercicio 197 sobre la relación entre el tipo de carácter social de los conductores y el número de sanciones por infracciones de circulación sufridas, se pide:

1. *Formular la hipótesis nula.*
2. *Contrastarla empleando el test pertinente.*
3. *Interpretar los resultados.*

Fórmula a aplicar:

$$F = \frac{\eta^2}{1 - \eta^2} \cdot \frac{N - K}{K - 1}$$

Distribución de F

$p = .05$

n_2 \ n_1	1	2	3	4	5	6	8	12	24	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	238.9	243.9	249.0	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.41	19.45	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.84	8.74	8.64	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.91	5.77	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.68	4.53	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.00	3.84	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.57	3.41	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.28	3.12	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.07	2.90	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.91	2.74	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.79	2.61	2.40
12	4.75	3.88	3.49	3.26	3.11	3.00	2.85	2.69	2.50	2.30
13	4 67	3.80	3.41	3.18	3.02	2.92	2.77	2.60	2.42	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.53	2.35	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.48	2.29	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.42	2.24	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.38	2.19	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.34	2.15	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.31	2.11	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.28	2.08	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.25	2.05	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.23	2.03	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.38	2.20	2.00	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.18	1.98	1.73
25	4.24	3.38	2.99	2.76	2.60	2.49	2.34	2.16	1.96	1.71
26	4.22	3.37	2.98	2.74	2.59	2.47	2.32	2.15	1.95	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.30	2.13	1.93	1.67
28	4.20	3.34	2.95	2.71	2.56	2.44	2.29	2.12	1.91	1.65
29	4.18	3.33	2.93	2.70	2.54	2.43	2.28	2.10	1.90	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.09	1.89	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.00	1.79	1.51
60	4.00	3.15	2.76	2.52	2.37	2.25	2.10	1.92	1.70	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.02	1.83	1.61	1.25
∞	3.84	2.99	2.60	2.37	2.21	2.09	1.94	1.75	1.52	1.00

Los valores de n_1 y n_2 representan los grados de libertad asociados a las estimaciones mayores y menores respectivamente de la variancia.

Distribución de F

$p = .01$ (continuación)

n_2 \ n_1	1	2	3	4	5	6	8	12	24	∞
1	4052	4999	5403	5625	5764	5859	5981	6106	6234	6366
2	98.49	99.01	99.17	99.25	99.30	99.33	99.36	99.42	99.46	99.50
3	34.12	30.81	29.46	28.71	28.24	27.91	27.49	27.05	26.60	26.12
4	21.20	18.00	16.69	15.98	15.52	15.21	14.80	14.37	13.93	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.27	9.89	9.47	9.02
6	13.74	10.92	9.78	9.15	8.75	8.47	8.10	7.72	7.31	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.84	6.47	6.07	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.03	5.67	5.28	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.47	5.11	4.73	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.06	.71	4.33	3.91
11	9.65	7.20	6.22	5.67	5.32	5.07	4.74	4.40	4.02	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.50	4.16	3.78	3.36
13	9.07	6.70	5.74	5.20	4.86	4.62	4.30	3.96	3.59	3.16
14	8.86	6.51	5.56	5.03	4.69	4.46	4.14	3.80	3.43	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.00	3.67	3.29	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	3.89	3.55	3.18	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.79	3.45	3.08	2.65
18	8.28	6.01	5.09	4.58	4.25	4.01	3.71	3.37	3.00	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.63	3.30	2.92	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.56	3.23	2.86	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.51	3.17	2.80	2.36
22	7.94	5.72	4.82	4.31	3.99	3.76	3.45	3.12	2.75	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.41	3.07	2.70	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.36	3.03	2.66	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.32	2.99	2.62	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.29	2.96	2.58	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.26	2.93	2.55	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.23	2.90	2.52	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.20	2.87	2.49	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.17	2.84	2.47	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	2.99	2.66	2.29	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.82	2.50	2.12	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.66	2.34	1.95	1.38
∞	6.64	4.60	3.78	3.32	3.02	2.80	2.51	2.18	1.79	1.00

Los valores de n_1 y n_2 representan los grados de libertad asociados a las estimaciones mayores y menores respectivamente de la variancia.

RESPUESTA

1. Hipótesis nula: No hay asociación entre el tipo de temperamento y el número de sanciones sufridas.

2. Contraste: Aplicando la fórmula se tiene:

$$F = \frac{0{,}24}{1 - 0{,}24} \times \frac{400 - 3}{3 - 1} = 0{,}31 \times 198{,}50 = 61{,}53$$

Valor de F en las tablas para 397 grados de libertad (400 – 3) y 2 grados de libertad (3 – 1), respectivamente, 26,14 aproximadamente, a un nivel de significación de 0,01.

Interpretación: El valor de F obtenido en el ejercicio es muy superior, luego se rechaza la hipótesis nula. Existe también en la población una relación entre el tipo de temperamento y las sanciones sufridas, y la asociación que pone de manifiesto la muestra no es debida al azar.

211

La aplicación de la escala de Chapin para medir el status socioeconómico a dos muestras distintas de seis familias ha dado los siguientes resultados:

	PUNTUACIONES	
Familias	*Muestra A*	*Muestra B*
1	50	65
2	40	35
3	75	100
4	20	25
5	15	30
6	70	75

Se pide:

1. Formular la hipótesis nula.

2. Comprobarla por el método t para grupos pequeños, teniendo en cuenta que la fórmula a aplicar es la siguiente:

$$t = \frac{x_{od}}{\epsilon x_1 - x_2}, X$$

en la que:

x_{od} = *Diferencia de las medias experimentales obtenidas.*

$\epsilon x_1 - x_2$ = *Error de las diferencias de las medias.*

Este error en el caso, como en el presente, en que los individuos de un grupo están conectados a pares con los del otro, es:

$$\frac{\sigma_d}{\sqrt{N}}$$

en la que:

σ_d = *Desviación standard de la distribución de las diferencias entre los individuos apareados.*

N = *Número de apareamientos.*

RESPUESTA

Disposición de los cálculos:

Muestra 1	*Muestra 2*	*Diferencia*	*Media de las diferencias*	*Desviación*	*Desviación al cuadrado*
65	50	15	10	5	25
35	40	– 5	10	– 15	225
100	75	25	10	15	225
25	20	5	10	– 5	25
30	15	15	10	5	25
75	70	5	10	– 5	25

$$\sigma \sqrt{\frac{550}{6}} = \sqrt{91{,}66} = 9{,}57$$

$$\epsilon x_1 - x_2 = \frac{9{,}57}{\sqrt{N}} = \frac{9{,}57}{\sqrt{5}} = \frac{9{,}57}{2{,}33} = 4{,}10$$

Razón crítica:

$$t = \frac{45 - 55}{4{,}1} = \frac{10}{4{,}1} = 2{,}43$$

Hipótesis nula: Los dos grupos son iguales. Las diferencias entre ellos son debidas al azar.

Contraste: Grado de libertad: N – 1 = 5.

Valor de t en la tabla para 5, grado de libertad a un nivel de significación del 5 por 100, 2.751. Es mayor que 2,43, luego no se puede rechazar la hipótesis nula.

212

En una oficina de 110 empleados, el número medio de los días que faltaron al trabajo en el año X fue de 6.

Al año siguiente en una nuestra al azar de 9 empleados resultó una media de 8 días de enfermedad con una desviación típica de 5 días.

Se trata de saber, con base en esta muestra, estadísticamente, si al 0,5 de nivel de significación, el número medio de días de enfermedad para toda la oficina ha cambiado.

1. *Formular la hipótesis nula.*
2. *Realizar su contraste mediante el test apropiado.*
3. *Interpretar los resultados.*

RESPUESTA

1. H_0 = No hay cambio en la media de días de enfermedad de un año a otro.

2. El test apropiado es t por tratarse de una muestra pequeña a estos efectos $N < 120$ y no conocerse la desviación típica de la población.

Fórmulas:

$$t = x - a/s_x \; ; \quad S_x = \sqrt{\frac{s^2}{n} \cdot \frac{N-n}{N-1}} \; ;$$

$$S_x = \sqrt{\frac{25}{9} \cdot \frac{110-9}{110-1}} = 1{,}60; \quad t = \frac{8-6}{1{,}60} = 1{,}25$$

3. t en las tablas al nivel de significación de 0,5 es 2.306. El t obtenido es menor, 1,25, luego no se puede rechazar H_0.

De la muestra utilizada no se puede inferir una diferencia entre ambos años. (Basado en Ackoff, 1973, p. 188).

TABLA DE VALORES DE *t*

Grados de libertad	*5 por 100*	*1 por 100*	*Grados de libertad*	*5 por 100*	*1 por 100*
1	12.706	63.657	32	2.037	2.739
2	4.303	9,925	34	2.032	2.728
3	3.182	5.841	36	2.027	2.718
4	2.776	4.604	38	2.025	2.711
5	2.571	4.032	40	2.021	2.704
6	2.447	3.707	42	2.017	2.696
7	2.365	3.499	44	2.015	2.691
8	2.306	3.355	46	2.012	2.685
9	2.262	3.250	48	2.010	2.681
10	2.228	3.169	50	2.008	2.678
11	2.201	3.106	55	2.005	2.668
12	2.179	3.055	60	2.000	2.660
13	2.160	3.012	65	1.998	2.653
14	2.145	2.977	70	1.994	2.648
15	2.131	2.947	80	1.990	2.638
16	2.120	2.921	90	1.987	2.632
17	2.110	2.898	100	1.984	2.626
18	2.101	2.878	125	1.979	2.616
19	2.093	2.861	150	1.976	2.609
20	2.086	2.845	200	1.972	2.601
21	2.080	2.831	300	1.968	2.592
22	2.074	2.819	400	1.966	2.588
23	2.069	2.807	500	1.965	2.586
24	2.064	2.797	1.000	1.962	2.581
25	2.060	2.787	∞	1.960	2.576
26	2.056	2.779			
27	2.052	2.771			
28	2.048	2.763			
29	2.045	2.756			
30	2.042	2.750			

213

En una pequeña ciudad se quiere contrastar la hipótesis, a un nivel de significación del 0,01, que en las próximas elecciones municipales, el presupuesto escolar propuesto tendrá mayoría de votos, es decir, por lo menos el 50 % .

Elegida una muestra al azar de N = 100 votantes, el 42 % mostraron su aprobación al presupuesto.

1. Formular la H_0.

2. Realizar su contraste mediante el test más apropiado.

3. Interpretar los resultados.

RESPUESTA

1. $H_0 = 50\ \%$ La diferencia entre la muestra y la población no es significativa, sino debida al azar.

2. El test apropiado es Z de la curva normal. Es muestra grande, ya que $p \cdot n$ ó $n \cdot (1 - p)$ es mayor que 5.

Aplicando la fórmula:

$$z = \frac{p - a}{\sqrt{\frac{a\,(1 - a)}{n}}}$$

se tiene:

$$z = \frac{42 - 50}{\sqrt{\frac{50 \cdot (50)}{100}}} = -1{,}60$$

3. El valor de z en la curva normal al 0,001 es de 2,58. Como 1,60 es inferior no se puede rechazar la H_0 con base en la muestra, ni se puede afirmar que el presupuesto de educación será aprobado por el 50 % de los electores, al menos. (Basado en Glass, 1973. pág. 321 y ss.)

214

En un colegio se escogieron al azar dos grupos de 25 estudiantes para hacer un experimento. Al primer grupo se le enseñó una materia aplicando un método nuevo y al segundo aplicando el método antiguo.

Los resultados obtenidos fueron los siguientes:

Grupo 1	*Grupo 2*
$n_1 = 25$	$n_2 = 25$
$\bar{x}_1 = 7{,}65$	$\bar{x}_2 = 6{,}00$
$s_1 = 6{,}5$	$s_2 = 5{,}90$

1. Formular la H_0.

2. Aplicar el test apropiado, hacer su contraste al nivel de significación de 0,5.

3. Interpretar los resultados.

RESPUESTA

1. $H_0 = \bar{x}_1 = \bar{x}_2$: no hay diferencia en ambos grupos en la media de respuestas correctas.

2. El test apropiado es el t. Al tratarse de una muestra pequeña, hay que aplicar la fórmula siguiente:

$$t = \frac{x_1 - x_2}{\sqrt{\frac{n_1 s_1^{\,2} + n_2 s_2^{\,2}}{n_1 + n_2 - 2} \left(\frac{n_1 + n_2}{n_1 \cdot n_2}\right)}}$$

sustituyendo se tiene:

$$t = \frac{7{,}65 - 6}{\sqrt{\frac{25 \cdot 6{,}5 + 25 \cdot 5{,}9}{25 + 25 - 2} \cdot \frac{25 + 25}{25 \cdot 25}}} =$$

$$= \frac{1{,}65}{2{,}53 \times (28)} = 2{,}32$$

El valor crítico de t para 48 grados de libertad es de 2,01.

El t obtenido es mayor, luego se rechaza la hipótesis nula. Es significativa la diferencia entre las medias de los dos grupos.

215

En un experimento realizado con dos grupos de enseñanza media, uno de 50 estudiantes y otro de 50, ha resultado que después de un cursillo sobre los efectos de las drogas, dado al grupo experimental, un 25 % de éste se declaraba favorablemente a la experiencia de probar la heroina y un 40 % en el grupo de control.

1. Formular la H_0.

2. Aplicar el test apropiado para un contraste a nivel de 0,5.

3. Interpretar los resultados.

RESPUESTA

1. $H_0 = P_1 = P_2$. No hay diferencia en ambos grupos en el porcentaje indicado.

2. El test apropiado es Z y la fórmula al tratarse de una muestra pequeña es:

$$z = \frac{P_1 - P_2}{\sqrt{\left(\frac{P_1 + P_2}{n_1 + n_2}\right) \cdot \left(1 - \frac{(P_1) + P_2}{n_1 + n_2}\right) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Sustituyendo en la fórmula se tiene:

$$z = \frac{40 - 25}{\sqrt{\left(\frac{25+40}{50+50}\right) \cdot \left(1 - \frac{25+40}{50+50}\right) \cdot \left(\frac{1}{50} + \frac{1}{50}\right)}} =$$

$$= \frac{15}{\sqrt{1{,}44 - 0{,}44 \cdot (0{,}09)}} = 0{,}15/0{,}84 = 1{,}76\ \sigma$$

El valor de Z para el nivel de significación de 0,5 es 1,96.

El Z obtenido es menor. No se puede rechazar la H_0. No hay diferencias significativas entre los dos grupos.

216

Antes y después de una conferencia sobre la eutanasia se pidió a un grupo de 100 personas que manifestaran si la aprobaban o la rechazaban. La tabla de rotación de los resultados fué la siguiente:

Después de la conferencia	*Antes de la conferencia*		
	Aprueban	*Desaprueban*	*Total*
Aprueban	15	45	60
Desaprueban	15	25	40
Total	30	70	100

1. Formular la H_0.
2. Contrastar con el test adecuado a un nivel de significación del 0,05.
3. Interpretar los resultados.

RESPUESTA

1. H_0 = No hay variación en el grupo de antes a después de la conferencia. Las diferencias observadas son explicables por el azar.

2. El test adecuado es Z para muestras dependientes. La fórmula es la siguiente:

$$z = \frac{d - a}{\sqrt{d + a}}; \quad \begin{array}{|c|c|}\hline a & b \\ \hline c & d \\ \hline \end{array}; \quad z = \frac{25 - 15}{\sqrt{25 + 15}} = 6{,}32\ \sigma$$

3. El valor crítico de Z para un nivel de significación del 0,05 es 1,96.

La Z obtenida es mucho mayor. Luego se rechaza la H_0.

La variación en el grupo es estadísticamente significativa a un nivel de probabilidad del 5 % (Basado en Glass, 1973, pág. 327).

28. TESTS NO PARAMETRICOS

En el capítulo anterior de esta obra se hace alusión a estos test estadísticos no paramétricos, pero únicamente se estudia una prueba que puede considerarse como tal, la de ji cuadrado.

En general se pueden caracterizar los tests no paramétricos como aquellos que conservando la finalidad básica de los tests, de servir para contrastar la validez de la generalización de los resultados obtenidos de una muestra a toda la población de la que se ha obtenido la muestra, no exigen, como los test paramétricos, el cumplimiento en los datos del requisito de la distribución normal de la variable o variables observadas en la población investigada.

Como puede ocurrir que no se cumplan a veces en las investigaciones sociales las condiciones exigidas para la aplicación de los test paramétricos, el conocimiento y empleo de estos test no paramétricos puede ser útil para los investigadores sociales en muchos casos.

Entre las ventajas de estos test se pueden señalar con S. Siegel (1975, p. 53) las siguientes:

1. La primera es la indicada, es decir, que no dependen de la forma normal de la distribución de la variable estudiada en la población de la que se tomó la muestra.

2. Son especialmente útiles en el caso de muestras pequeñas ya que en las muestras grandes se puede prescindir, de acuerdo con la ley de los grandes números, del supuesto de normalidad. Si la muestra es tan pequeña como n = 6, entonces los test no paramétricos son la única alternativa posible a no ser que se conozca la normalidad exacta de la población.

3. Los test paramétricos no requieren que se llegue a un nivel de intervalo en las medidas de las variables. En ellos es suficiente el nivel de medida ordinal y se pueden aplicar también a veces a datos clasificatorios de nivel de medida simplemente nominal.

4. Estas pruebas son más fáciles de aplicar que las no paramétricas.

La desventaja de los tests no paramétricos consiste fundamentalmente en que, al basarse en supuestos más débiles que los no paramétricos, no son tan potentes como éstos, en el sentido de que pueden disipar o no aprovechar debidamente la información proporcionada por los datos y comportar un riesgo mayor de error de tipo I.

El grado de esta disipación o pérdida de información se determina hallando la potencia-eficiencia de la prueba o eficiencia de la fuerza. Como escribe Blalock (1964, p. 214) se emplea comúnmente el término *eficiencia de la fuerza* para designar la fuerza de una prueba determinada en relación con su alternativa más fuerte. Si designamos la eficiencia de la fuerza de una de dichas pruebas no paramétricas como el 95%, queremos decir que la fuerza de la prueba no paramétrica sirviéndonos de 100 casos es igual a la del test t sirviéndonos de 95 casos, *si el modelo empleado en la prueba t es correcto.*

Por tanto, este defecto de potencia de las pruebas no paramétricas con relación a las paramétricas se puede compensar aumentando el tamaño de la muestra en la proporción debida para que su potencial-eficiencia llegue al 100%, p. ej. si es del 29 % en un 5 % más.

Existe un gran número de test estadísticos no paramétricos y si bien muchos de ellos tienen aplicaciones similares y frecuentemente son intercambiables, no dejan de presentar sus peculiaridades propias.

Aquí sólo se presentan algunos de los existentes, de aplicación más generalizada en su campo.

Para obtener una información bastante completa sobre este tema se puede consultar la obra de S. Siegel citada, *Estadística no paramétrica aplicada a las ciencias de la conducta,* (Ed. Trillas, 1964).

PRUEBA DE LAS SECUENCIAS O RACHAS DE WALD-WOLLFOWITZ

Aplicaciones.– Este test sirve para contrastar dos muestras entre sí y decidir si las poblaciones de las que se han obtenido las muestras son iguales en tendencia central, dispersión, etc. respecto a la característica o variable investigada.

Supuestos.— Esta prueba exige que las muestras sean independientes, que la variable tenga una distribución continua y que el nivel de medida de la misma logrado sea ordinal.

Operaciones que comprende.— En síntesis (Vid. Siegel 1975, p. 173 son las siguientes):

1. Se colocan los valores o puntuaciones de las dos muestras en una sola serie ordenada.

2. Se determina el número de secuencias o rachas. Por secuencia o racha se entiende cada sucesión continua de datos de la misma muestra o grupo.

Para ello se puede ordenar la serie de todos los rangos de menor a mayor y trazar unas líneas contínuas debajo de los rangos seguidos que pertenecen a la primera muestra, y una línea encima de los rangos seguidos que pertenecen a la segunda.

Si las muestras son pequeñas, es decir, en este caso, si tanto n_1 como n_2 son 20 o menos, entonces el número de secuencias o rachas se contrasta con la tabla que da los valores críticos de las rachas al nivel de significación del 5%. Adviértase que aquí si el valor observado de r, rachas, es igual o menor que el crítico se rechaza la hipótesis nula.

4. Si las muestras son grandes, mayores de 20, el contraste se hace con la curva normal, obteniendo los valores z, mediante la siguiente fórmula:

$$z = \frac{\left| r - \left(\frac{2\, n_1\, n_2}{n_1 + n_2} + 1 \right) \right| - .5}{\sqrt{\frac{2\, n_1\, n_2\, (2\, n_1\, n_2 - n_1 - n_2)}{(n_1 + n_2)^2\, (n_1 + n_2 - 1)}}}$$

(Ver ejercicios 217 y 218)

PRUEBA DE MAC NEMAR PARA LA SIGNIFICACION DE LOS CAMBIOS

Función.— Esta prueba es útil para contrastar la significación de los cambios producidos en un grupo dicotomizado, en relación a alguna variable, de un tiempo t_1 a t_2.

Por tanto, es aplicable en los estudios llamados de panel y en los diseños experimentales con un solo grupo que se somete a observación antes y después de haber sufrido el impacto de una variable, que puede actuar también espontáneamente.

Presupuestos.— Exige, pues dos observaciones del mismo grupo del tipo antes-después y un nivel de medida nominal, clasificatorio, o bien ordinal del tipo, más-menos, si bien la variable puede ser en realidad continua, aunque sólo se logre un nivel de medida dicotómico en ella.

En este último caso hay que hacer la llamada corrección de continuidad en la fórmula.

Procedimiento:

1. Este procedimiento se basa en la aplicación de la fórmula de χ^2 en relación únicamente a las casillas de la tabla de 2 x 2 que representen cambios.

2. Por ello sólo se puede aplicar si las frecuencias esperadas de los cambios en A y D son superiores a 5.

3. Supuesto lo anterior, la fórmula de χ^2 es la siguiente para un grado de libertad:

$$\chi^2 = \frac{(B - C)^2}{B + C}$$

A	B
C	D

Con relación por continuidad, la fórmula se convierte en la siguiente:

$$\chi^2 = (|B - C| - 1)^2 / B + C$$

4. El contraste se hace por la tabla de χ^2. Si la prueba es unilateral se divide el valor crítico de la tabla por 2. (Ver ejercicio 221).

PRUEBA DE KOLMOGOROV-SMIRNOV

Fundamento.— Es muy simple. Se basa en que si dos grupos o muestras son iguales en relación a la variable investigada no debe haber diferencia entre ellas en sus frecuencias acumuladas.

Aplicación y supuestos.— Son, respecto a la comparación de dos muestras, iguales a las de la prueba anterior de las rachas o secuencias. También se puede aplicar cuando la muestra es sólo una y no dos. Esta prueba tiene una potencia-eficiencia alta. Del 96% en relación a la prueba t.

Operaciones que comprende:

1. Se forma la distribución de frecuencias acumuladas de los dos grupos en números enteros y en proporciones, debiendo tener ambas distribuciones el mismo número de intervalos.

2. Se hallan las diferencias entre las frecuencias relativas o entre las proporciones acumuladas de los dos grupos.

3. Por simple inspección se busca la mayor de estas diferencias, a la que se da el nombre de D.

4. Para determinar la significación de D hay que distinguir según que las muestras sean pequeñas ≤ 40, o grandes, > 40, y según que la prueba en éstas últimas sea de una o de dos colas.

5. Cuando el n de los dos grupos es igual o menor a 40 , se usa la tabla que da los valores críticos según distintos niveles de significación para las pruebas de una y de dos colas.

6. Cuando n_1 y n_2 son mayores que 40 y la prueba de dos colas se utiliza la tabla, previa aplicación de la fórmula siguiente para el nivel de significación del 5 :

$$136\sqrt{\frac{n_1 + n_2}{n_1\ n_2}}$$

En este caso no es necesario que n_1 sea igual a n_2 .

7. Si las muestras son grandes y la prueba es de una sola cola se utiliza como contraste la distribución muestral de χ^2 . Para obtener el valor empírico de χ^2 se utiliza la siguiente fórmula:

$$\chi^2 = 4\ D^2\ \frac{n_1\ n_2}{n_1 + n_2}$$

(Ver ejercicios 219 y 220)

PRUEBA DE LOS SIGNOS

Función.— Esta prueba es aplicable al mismo tipo de estudios de panel (antes y después) que la anterior. También vale para la prueba de las diferencias entre muestras dependientes, relacionadas entre sí, como de padres e hijos, maridos y mujeres etc.

Presupuestos.— Medidas pareadas de nivel ordinal o que supongan un cierto orden, y que pueda considerarse como continua la variación en la realidad de la variable considerada.

Procedimiento:

1. Se halla el signo de la diferencia entre los dos elementos de cada pareja.

2. Se halla el número de parejas con diferencias en los signos.

3. Se determina X, que es igual al número de signos. (Número menor entre los +, - y =).

4. Si n es menor que 25 el X obtenido se contrasta con el p crítico de la tabla, en el caso de prueba de una sola cola.

Si la prueba es de dos colas se duplica el valor de p que figura en la tabla D.

5. Si n es mayor que 25 el contraste se realiza mediante la tabla de áreas de la curva normal.

6. La fórmula para calcular z es la siguiente: (Ver ejercicios 222 y 223)

$$z = \frac{X - 1/2\ N}{1/2\sqrt{N}}$$

ANALISIS DE VARIANZA NO PARAMETRICO: LA PRUEBA DE KRUSKAL Y WALLIS

El análisis de varianza, como prueba paramétrica que es, exige, para que se pueda aplicar, que se cumplan en el caso analizado las condiciones o supuestos aludidos siguientes:

1. Que las variables tengan un nivel de medida por lo menos escala de intervalo;

2. Que las muestras o grupos analizados sean independientes y hayan sido obtenidos de poblaciones distribuídas normalmente;

3. Que estas poblaciones tengan la misma varianza;

4. Que no exista interacción.

Generalmente, si no hay evidencia contraria, se supone que se dan dichas condiciones, de modo que, como señala Siegel (1975, p. 40) cuando hay razones para creer que las mismas se cumplen en el caso examinado, se debe escoger el análisis de varianza por ser una prueba más potente y eficaz que las no parámetricas.

Pero cuando no se cumplen los requisitos indicados no se debe emplear el análisis de varianza y hay que utilizar en su lugar pruebas no paramétricas, así llamadas porque no están sujetas a dichas condiciones.

Una de ellas es la de Kruskal y Wallis, que es aplicable para realizar el análisis de varianza no paramétrico cuando hay en el modelo una sola variable nominal.

Esta prueba exige que las muestras sean independientes y que el nivel de medida de las variables sea ordinal. En ella todas las puntuaciones o valores de las variables se sustituyen por rangos correlativos, empezando

por la puntuación más baja a la que se da el rango 1, el dos a la siguiente y así sucesivamente.

La fórmula de esta prueba es la siguiente:

$$H = \frac{\frac{12}{N(N+1)} \Sigma \frac{R_i^2}{n_i} - 3(N+1)}{1 - \Sigma T/(N^3 - N)}$$

en la que
k = Número de muestras o grupos,
n_i = Número de casos de cada muestra o grupo i,
N = Número total, o suma de todos los casos de las muestras,
R_i = Suma de rangos en la muestra o grupo i.

La distribución del estadístico H de Kruskal y Wallis es aproximadamente la de χ^2 Por ello, para contrastar en esta prueba la hipótesis nula se utiliza la tabla de valores críticos de la distribución χ^2

Si existen repeticiones de rangos o empates se da a todos ellos el valor medio que correspondería a los rangos si no fueran iguales, y se corrige la fórmula dividiendo su resultado,

$$1 - \Sigma T/(N^3 - N)$$

en la que $T = t^3 - t$
siendo t el número de rangos ligados o repetidos.
(Ver ejercicio 224)

217 EJERCICIOS

Se aplicó a dos grupos de varones mayores de 23 años formado el primero por 12 casados y el segundo por doce solteros, una escala de feminismo-machismo.

Las puntuaciones fueron las siguientes:

Casados	*Solteros*
81	56
70	42
75	59
65	15
110	10
62	9
115	17
50	28
140	35
43	22
58	16

A) Indicar las variables y su tipo y la prueba que podría ser adecuada para decidir estadísticamente si las diferencias en feminismo entre los dos grupos son significativas.

RESPUESTA

La variable básica del ejercicio es el feminismo con nivel de medida ordinal pero que, en sí, si hubiera instrumentos de medida adecuados, puede considerarse como una variable continua.

En segundo lugar, los datos corresponden a dos grupos distintos que, cuando uno de ellos está formado por casados y el otro por solteros, constituyen dos categorías de la variable nominal, estado civil.

Al ser las muestras independientes, el nivel de medida ordinal y continua la distribución de la variable en la población y dado que no se dice se quiera investigar una diferencia específica entre los grupos (p. ej. en la media etc.), puede emplearse la prueba de las secuencias o rachas de Wald-Wolffowitz.

B) Formular la H_0 y contrastarla mediante el test indicado.

RESPUESTA

H_0: No hay diferencia entre casados y solteros en feminismo. Las diferencias observadas se explican por la variación muestral.

Contraste.– Al tratarse de muestras pequeñas, menos de 20, r, el número de rachas o secuencias, se halla disponiendo por orden en una sola serie los dos grupos de puntuaciones y marcando las rachas de puntuaciones de casados con una línea continua debajo, y las de los solteros con otra línea encima. Tendríamos pues:

9	10	15	16	17	20	22	28	35	42	43	50	56
58	59	62	65	70	75	81	100	110	115	140		

Hay seis rachas o secuencias, por tanto, r = 6.

El r crítico en la tabla que figura en la página 404 es 7. Todos los resultados iguales o menores al punto crítico son significativos al nivel del 5%. El obtenido es menor, luego es significativo y se puede rechazar la H_0 con una probabilidad de error del 5%. Las dos poblaciones de casados y solteros de donde se han obtenido las muestras observadas son distintas en cuanto a machismo-feminismo.

218

Supuesto que los dos grupos del ejercicio anterior están formados, en lugar de por 12 y 12, uno por 30 casados y el otro por 25 solteros y que las rachas obtenidas por la ordenación en la forma indicada de las puntuaciones son nueve, se pide indicar qué habría que hacer para contrastar la H_0 mediante el test de las rachas de Wald-Wolfowitz y efectuar el contraste.

RESPUESTA

Al tener un grupo 30 personas y el otro 25 se trata de muestras grandes y, por tanto, para realizar el contraste hay que aplicar la fórmula pertinente siguiente:

$$z = \frac{r - \frac{2 n_1 n_2}{n_1 + n_2} + 1}{\sqrt{\frac{2 n_1 n_2 (2 n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}}}$$

$$z = \frac{9 - \frac{2(25)30}{25 + 30} + 1}{\sqrt{\frac{[2(25)(30)][2(25)30 - 25 - 30]}{(25 + 30)^2 (25 + 30 - 1)}}} = 4.82$$

Una z de 4.82 es significativa a un nivel mucho más elevado del 5%. Luego se rechaza la H_0, y se puede concluir que en las poblaciones investigadas existen diferencias significativas entre casados y solteros en cuanto al grado de machismo-feminismo.

219

Dadas las puntuaciones en la escalda de actitudes de machismo-feminismo de los dos grupos de casados y solteros del ejercicio número 41, efectuar el contraste de la H_0 utilizando la técnica de Kolmogorov-Smirnov.

RESPUESTA

Se trata de dos muestras pequeñas, luego hay que utilizar la primera modalidad indicada de este test.

Para ello como el recorrido de las dos distribuciones va desde 9 hasta 140, se pueden formar 7 intervalos de 20 unidades cada uno y consignar separadamente los casos de cada distribución que caen dentro de cada intervalo, obteniendo después las frecuencias relativas.

Es conveniente formar para ello la siguiente tabla:

Intervalos	*Frecuencias*		*Frecuencia relativa*		*Diferencia*
	Casados	*Solteros*	*Casados*	*Solteros*	
1–29	0	8	0	8/12	8/12
30–50	2	2	2/12	10/12	8/12
51–71	4	2	6/12	12/12	8/12
72–92	2	0	8/12	12/12	4/12
93–113	2	0	10/12	12/12	2/12
114–134	1	0	11/12	12/12	1/12
135–155	1	0	12/12	12/12	0

D es el numerador de la diferencia mayor.

$$D = 8/12 = 8.$$

Valor crítico de D para un n de 12 y a un nivel de significación del 5 % = 6.

El D observado es mayor; luego se rechaza la hipótesis nula y se llega a la misma conclusión que con la prueba de las rachas. Utilizar la tabla de la pág. 480.

220

Dada la siguiente tabla:

Clase social	*Aspiraciones*	
	Bajas	*Altas*
Inferior-inferior	58	31
Inferior-superior	51	46
Media-inferior	47	53
Media-superior	44	73
Superior-inferior	22	51
Superior-superior	14	20
	236	274

Fuente: Blalock 1964, p. 227.

se quiere contrastar la hipótesis de que a una posición más elevada en la clase social las aspiraciones serán más elevadas.

A) Indicar las variables de la tabla y el test no paramétrico que se podría utilizar para contrastar la H_0 derivada de la hipótesis de investigación señalada, teniendo en cuenta que los datos proceden de una muestra total obtenida al azar.

RESPUESTA

Las variables son dos: las aspiraciones (dicotomizada) altas y bajas y la clase social (ordinal). La primera sería dependiente y la segunda independiente.

Las dos categorías de la variable aspiraciones se pueden considerar como muestras aleatorias independientes de las poblaciones más amplias de varones adultos con aspiraciones altas y bajas respectivamente, ya que, como dice Blaclock, una muestra total completamente al azar asegura la independencia entre las dos submuestras que se pueden escoger.

Dado que, además, el nivel de medida es ordinal, que se puede estimar que la distribución de la variable aspiraciones tiene una distribución continua y que el nivel de medida logrado es ordinal, se dan los supuestos para aplicar la prueba de Kolmogorov-Smirnov de una sola cola al estar determinada la variable independiente, la clase social.

B) Efectuar el contraste mediante el test acabado de indicar.

RESPUESTA

H_0: No hay diferencias en las aspiraciones según la clase social. Las variaciones observadas son debidas al azar.

Contraste.– Al ser la muestra grande, mayor de 40 y el test unilateral, hay que hallar D, la diferencia mayor entre las frecuencias relativas acumuladas de las dos distribuciones, luego aplicar esta fórmula:

$$\chi^2 = 4 D^2 \frac{n_1 n_2}{n_1 + n_2}$$

Conviene formar esta tabla:

Clase social	*Aspiraciones de cambios*				*Diferencia*
	Altas, F		*Bajas, F'*		*F - F'*
Debajo de la inferior-superior	58	.246	31	.113	.133
Debajo de la media-inferior	109	.462	77	.281	.181
Debajo de la media-superior	156	.661	130	.474	.187
Debajo de la superior-inferior	200	.847	203	.741	.106
Debajo de la superior-superior	222	.941	254	.927	.014
Total	236	1.000	274	1.000	

D, la mayor diferencia de frecuencias, es .187, por tanto:

$$\chi^2 = 4\,(.187)^2\,\frac{236\,(274)}{236+274} = 17.75$$

El χ^2 crítico en la tabla para dos grados de libertad es 13.82 al nivel de significación del 1 por mil.

El χ^2 observado es mayor, luego se puede rechazar la H_0 con una probabilidad de error del 1 por mil. Hay diferencias muy significativas en las aspiraciones según la clase social.

221

Se eligió aleatoriamente un grupo de 50 personas pertenecientes a una comunidad determinada. Todas ellas siguieron un cursillo sobre la energía nuclear. Antes y después del cursillo se midió la aceptación de la energía nuclear por el grupo mediante un cuestionario. Los resultados de las dos observaciones se reflejan en la siguiente tabla:

Antes del cursillo	*Después del cursillo*		
	Aceptación	*Rechazo*	*Total*
Aceptación	2	28	30
Rechazo	8	12	20
Total	10	40	50

A) Determinar las variables de la tabla y la clase de prueba que sería aplicable para ver si las diferencias entre antes y después del cursillo son significativas.

RESPUESTA

La variable es una, aceptación-rechazo de la energía nuclear, pero observada en dos tiempos distintos.

Por ello se puede utilizar la prueba de Mc Nemar, dado que se trata de un estudio de panel y la variable es dicotómica de aceptación-rechazo.

B) Formular la hipótesis nula, aplicar la prueba e interpretar los resultados.

RESPUESTA

H_0: No hay diferencia en el grupo sobre la aceptación de la energía nuclear antes y después del cursillo. La advertida se explica por la fluctuación de las muestras.

Contraste.– Dado que se puede admitir una continuidad real en la variable aceptación de la energía nuclear, se puede aplicar la fórmula corregida siguiente:

$$\chi^2 = \frac{[(B-C)-1]^2}{B+C} = \frac{(28-8-1)^2}{28+8} = 10.02$$

En la tabla da un valor de 6.64 a un nivel de significación del 1% . Por tanto se puede rechazar la H_0. Son significativas las diferencias observadas con una probabilidad de error del 1

222

Respecto a una muestra de doce parejas de hermanos, uno de ellos varón y el otro hembra, de edad entre los 13 a los 17 años, se midió mediante una escala la estimación que tenían de su madre obteniéndose las siguientes puntuaciones:

Parejas	*Estimación de la madre por el hermano*	
	Varón	*Hembra*
1	24	18
2	28	23
3	26	27
4	30	20
5	18	10
6	15	15
7	29	30
8	25	25
9	35	34
10	38	40
11	21	15
12	25	14

A) Determinar las variables de la tabla y el tipo de prueba adecuada para ver si la diferencia entre los dos grupos de hermanos son significativas.

RESPUESTA

La variable básica, dependiente, es el grado de estimación de la madre. Al ser uno de los grupos de varones y el otro de hembras, los dos constituyen la varible sexo, que se puede considerar como variable independiente.

Se puede utilizar la prueba de los signos para muestras pequeñas, al ser n menor que 25, tratarse de grupos de parejas de hermanos y poderse estimar como continua la variable estimación de la madre.

B) Formular la H_0, aplicar la prueba indicada e interpretar los resultados.

RESPUESTA

No hay diferencia entre los dos grupos o entre los hijos y las hijas en la estimación de la madre. Las diferencias observadas se pueden explicar por la variación muestral.

Contraste. –Conviene formar la siguiente tabla:

Parejas	*Puntuaciones en la estimación de la madre*		
	Hermano	*Hermana*	*Signo*
1	24	18	+
2	28	23	+
3	26	27	–
4	30	20	+
5	18	10	+
6	15	5	0
7	29	30	–
8	25	25	0
9	35	34	+
10	38	40	–
11	21	15	+
12	25	14	+

Hay siete signos positivos y tres negativos. Por tanto x es el menor número igual a 3. N, al existir dos empates, es 10.

El valor de p en la tabla para $X = 3$ y $n = 10$ es .172. Como la prueba es de una sola cola, al estar determinada la variable dependiente, el sexo, hay que multiplicar esta cifra por dos. Por tanto, $p = .344$.

Como es sabido se suele aceptar como significativa en los estudios sociales una p. o probabilidad máxima permitida de .05 generalmente. La p. obtenida es mucho mayor. Luego no se puede rechazar la H_0. No hay diferencias según el sexo de los hermanos en la estimación de la madre. Las diferencias observadas se explican por la variación de las muestras.

223

Dado el caso supuesto en el ejercicio número 310 de aplicación de la prueba de Mc Nemar efectuar su contraste mediante la utilización de la prueba de los signos.

RESPUESTA

Por ser n mayor que 25 es aplicable la prueba de los signos para muestras grandes.

En la tabla 2 x 2 se ve que 28 personas pasan de + (aceptación), en el tiempo 1, a – (rechazo) en el tiempo 2, y 8 de – en t_1 a + en t_2; por tanto x es 28 y $n = 36$.

Fórmula:

$$z = \frac{(x \pm .05) - 1/2\,N}{1/2\sqrt{N}} \qquad z = \frac{27.50 - 1/2\,(36)}{1/2\sqrt{36}} = 3.5$$

Una z de 3.5 es significativa a más del 1%. Luego se rechaza la H_0. El resultado concuerda con la prueba de Mc Nemar.

224

Dada la siguiente tabla:

Puntuaciones en una escala de actitudes sobre feminismo, tipo Likert, de tres grupos de universitarios:

Universitarios orientados hacia		
La enseñanza	*La Admon. Pública*	*La empresa privada*
31	21	16
12	32	36
29	31	19
28	21	21
15	20	21
	29	

se pide realizar el análisis de varianza utilizando la prueba no paramétrica de Kruskal y Wallis, supuesto que se dan las condiciones exigidas de constituir los grupos muestras independientes, y dado que el nivel logrado con las escalas de actitudes tipo Likert es el ordinal, y para ello:

A) Formular la hipótesis nula.

RESPUESTA

Los valores medios en feminismo de los tres grupos son iguales, de tal modo que la distinta orientación profesional de los universitarios no influye en su feminismo.

225

B) Calcular el valor de H del test de Kruskal y Wallis con los datos de la tabla.

RESPUESTA

La fórmula a aplicar es:

$$H = \frac{\frac{12}{N(N+1)} \Sigma \frac{R_j^2}{n_j} - 3(N+1)}{1 - \frac{\Sigma T}{N^3 - N}}$$

Por lo tanto los datos que se precisan son:

$$\Sigma R_j \qquad y \qquad \Sigma T$$

Tabla de la prueba

Enseñanza		*Admon. Púb.*		*Empresa Privada*	
Feminismo	*Rango* ΣR_j	*Feminismo*	*Rango* ΣR_j	*Feminismo*	*Rango* ΣR_j
31	16.5	21	17.5	16	3
12	1	32	15	36	16
29	11.5	32	13.5	19	4
28	10	21	7.5	21	7.5
15	2	20	5	21	7.5
		29	11.5		
	38		60		38

$\Sigma T = t^3 - t$

$t = N^o$ empates $= 8$

$\Sigma T = 8^3 - t = 504$

$$H = \frac{\frac{12}{(16+1)16} \quad \frac{38^2}{5} + \frac{60^2}{6} + \frac{38^2}{5} - 3(16+1)}{1 - \frac{504}{4080}} = 3.36$$

Decisión.– El χ^2 crítico en la tabla para dos grados de libertad es 5.991 al nivel de significación del .05 y 7.824 al nivel de .02. Por tanto no se puede rechazar la hipótesis nula a estos niveles de riesgo de error. No existe un influjo de la orientación profesional de los universitarios en el feminismo.

Adviertase que los datos utilizados en este ejercicio son los mismos que los empleados en el ejercicio 19 de análisis de varianza, en el que el resultado era también no significativo. El análisis de varianza es una prueba más potente que la prueba de Kruskal y Wallis, como en general los tests paramétricos son más potentes que los no paramétricos y poseen más fuerza discriminativa para rechazar la hipótesis nula.

TABLAS

Tabla de valores críticos de r_s, el coeficiente de correlación de rangos de Spearman

N	*Nivel de significación (prueba de una cola)*	
	.05	.01
4	1.000	
5	.900	1.000
6	.829	.943
7	.714	.893
8	.643	.833
9	.600	.783
10	.564	.746
12	.506	.712
14	.456	.645
16	.425	.601
18	.399	.564
20	.377	.534
22	.359	.508
24	.343	.485
26	.329	.465
28	.317	.448
30	.306	.432

Tabla de probabilidades asociadas con valores tan grandes como los valores observados de S en el coeficiente de correlación de rango de Kendall

S	*Valores de N*				*S*	*Valores de N*		
	4	5	8	9		6	7	10
0	.625	.592	.548	.540	1	.500	.500	.500
2	.375	.408	.452	.460	3	.360	.386	.431
4	.167	.242	.360	.381	5	.235	.281	.364
6	.042	.117	.274	.306	7	.136	.191	.300
8		.042	.199	.238	9	.068	.119	.242
10		.0083	.138	.179	11	.028	.068	.190
12			.098	.130	13	.0083	.035	.146
14			.054	.090	15	.0014	.015	.108
16			.031	.060	17		.0054	.078
18			.016	.038	19		.0014	.054
20			.0071	.022	21		.00020	.036
22			.0028	.012	23			.023
24			.00087	.0063	25			.014
26			.00019	.0029	27			.0083
28			.000025	.0012	29			.0046
30				.00043	31			.0023
32				.00012	33			.0011
34				.000025	35			.00047
36				.0000028	37			.00018
					39			.000058
					41			.000015
					43			.0000028
					45			.00000028

Tabla de valores críticos de s en el coeficiente de concordancia de Kendall*

k	*N*					*Valores adicionales para N = 3*	
	3 †	4	5	6	7	*k*	*s*
Valores al nivel de significación 0.05							
3			64.4	103.9	157.3	9	54.0
4		49.5	88.4	143.3	217.0	12	71.9
5		62.6	112.3	182.4	276.2	14	83.8
6		75.7	136.1	221.4	335.2	16	95.8
8	48.1	101.7	183.7	299.0	453.1	18	107.7
10	60.0	127.8	231.2	376.7	571.0		
15	89.8	192.9	349.8	570.5	864.9		
20	119.7	258.0	468.5	764.4	1,158.7		
Valores al nivel de significación 0.01							
3			75.6	122.8	185.6	9	75.9
4		61.4	109.3	176.2	265.0	12	103.5
5		80.5	142.8	229.4	343.8	14	121.9
6		99.5	176.1	282.4	422.6	16	140.2
8	66.8	137.4	242.7	388.3	579.9	18	158.6
10	85.1	175.3	309.1	494.0	737.0		
15	131.0	269.8	475.2	758.2	1,129.5		
20	177.0	364.2	641.2	1,022.2	1,521.9		

Nótese que los valores críticos adicionales de *s* para N = 3 se dan en la columna derecha de esta tabla.

Tabla de valores críticos de r en la prueba de rachas

Para la prueba de las rachas de una muestra, cualquier valor de r es igual a o menor que el que aparece en la tabla F_I, o igual o mayor que el que aparece en la tabla F_{II} es significativo en el nivel 0.05. Para la prueba de las rachas de dos muestras de Wald-Wolfwitz, cualquier valor de r igual o menor que el que aparece en la tabla es significativo en el nivel 0.05.

TABLA F_I

n_1 \ n_2	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2											2	2	2	2	2	2	2	2	2
3					2	2	2	2	2	2	2	2	2	3	3	3	3	3	3
4				2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4
5			2	2	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4
6		2	2	3	3	3	3	4	4	4	4	5	5	5	5	5	5	6	6
7		2	2	3	3	3	4	4	5	5	5	5	5	6	6	6	6	6	6
8		2	3	3	3	4	4	5	5	5	6	6	6	6	6	7	7	7	7
9		2	3	3	4	4	5	5	5	6	6	6	7	7	7	7	8	8	8
10		2	3	3	4	5	5	5	6	6	7	7	7	7	8	8	8	8	9
11		2	3	4	4	5	5	6	6	7	7	7	8	8	8	9	9	9	9
12	2	2	3	4	4	5	6	6	7	7	7	8	8	8	9	9	9	10	10
13	2	2	3	4	5	5	6	6	7	7	8	8	9	9	9	10	10	10	10
14	2	2	3	4	5	5	6	7	7	8	8	9	9	9	10	10	10	11	11
15	2	3	3	4	5	6	6	7	7	8	8	9	9	10	10	11	11	11	12
16	2	3	4	4	5	6	6	7	8	8	9	9	10	10	11	11	11	12	12
17	2	3	4	4	5	6	7	7	8	9	9	10	10	11	11	11	12	12	13
18	2	3	4	5	5	6	7	8	8	9	9	10	10	11	11	12	12	13	13
19	2	3	4	5	6	6	7	8	8	9	10	10	11	11	12	12	13	13	13
20	2	3	4	5	6	6	7	8	9	9	10	10	11	12	12	13	13	13	14

TABLA F_{II}

n_1 \ n_2	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2																			
3																			
4			9	9															
5		9	10	10	11	11													
6		9	10	11	12	12	13	13	13	13									
7			11	12	13	13	14	14	14	14	15	15	15						
8			11	12	13	14	14	15	15	16	16	16	16	17	17	17	17	17	
9				13	14	14	15	16	16	16	17	17	18	18	18	18	18	18	
10				13	14	15	16	16	17	17	18	18	18	19	19	19	20	20	
11				13	14	15	16	17	17	18	19	19	19	20	20	20	21	21	
12				13	14	16	16	17	18	19	19	20	20	21	21	21	22	22	
13					15	16	17	18	19	19	20	20	21	22	22	23	23	23	
14					15	16	17	18	19	20	20	21	22	22	23	23	23	24	
15					15	16	18	18	19	20	21	22	22	23	23	24	24	25	
16						17	18	19	20	21	21	22	23	23	24	25	25	25	
17						17	18	19	20	21	22	23	23	24	25	25	25	26	
18						17	18	19	20	21	22	23	24	25	25	26	26	27	
19						17	18	20	21	22	23	23	24	25	26	26	27	27	
20						17	18	20	21	22	23	24	25	25	26	27	27	28	

Tabla de valores críticos de D en la prueba de una muestra de Kolmogorov-Smirnov

Tamaño de muestra (N)	*Nivel de significación para D = máximo $\|F_0(X) - S_N(X)\|$*				
	.20	.15	.10	.05	.01
1	.900	.925	.950	.975	.995
2	.684	.762	.776	.842	.929
3	.565	.597	.642	.708	.828
4	.494	.525	.564	.624	.733
5	.446	.474	.510	.565	.669
6	.410	.436	.470	.521	.618
7	.381	.405	.438	.486	.577
8	.358	.381	.411	.457	.543
9	.339	.360	.388	.432	.514
10	.322	.342	.368	.410	.490
11	.307	.326	.352	.391	.468
12	.295	.313	.338	.375	.450
13	.284	.302	.325	.361	.433
14	.274	.292	.324	.349	.418
15	.266	.283	.304	.338	.404
16	.258	.274	.295	.328	.392
17	.250	.266	.286	.318	.381
18	.244	.259	.278	.309	.371
19	.237	.252	.272	.301	.363
20	.231	.246	.264	.294	.356
25	.21	.22	.24	.27	.32
30	.19	.20	.22	.24	.29
35	.18	.19	.21	.23	.27
Más de 35	$\frac{1.07}{\sqrt{N}}$	$\frac{1.14}{\sqrt{N}}$	$\frac{1.22}{\sqrt{N}}$	$\frac{1.36}{\sqrt{N}}$	$\frac{1.63}{\sqrt{N}}$

Tabla de valores críticos de K_D en la prueba de dos muestras de Kolmogorov-Smirnov

(muestras pequeñas)

N	*Prueba de una cola**		*Prueba de dos colas†*	
	$\alpha = .05$	$\alpha = .01$	$\alpha = .05$	$\alpha = .01$
3	3	—	—	—
4	4	—	4	—
5	4	5	5	5
6	5	6	5	6
7	5	6	6	6
8	5	6	6	7
9	6	7	6	7
10	6	7	7	8
11	6	8	7	8
12	6	8	7	8
13	7	8	7	9
14	7	8	8	9
15	7	9	8	9
16	7	9	8	10
17	8	9	8	10
18	8	10	9	10
19	8	10	9	10
20	8	10	9	11
21	8	10	9	11
22	9	11	9	11
23	9	11	10	11
24	9	11	10	12
25	9	11	10	12
26	9	11	10	12
27	9	12	10	12
28	10	12	11	13
29	10	12	11	13
30	10	12	11	13
35	11	13	12	
40	11	14	13	

Tabla de valores críticos de D en la prueba de dos muestras de Kolmogorov-Smirnov

(muestras grandes: prueba de dos colas)

Nivel de significación	*Valores de D tan grandes que requieren rechazar H_0 en el nivel de significación indicado, donde D = máximo* $\lvert S_{n1}(X) - S_{n2}(X)\rvert$
.10	$1.22\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$
.05	$1.36\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$
.025	$1.48\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$
.01	$1.63\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$
.005	$1.73\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$
.001	$1.95\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$

Tabla de probabilidades asociadas con valores tan pequeños como los valores observados de x en la prueba binomial.

En el cuerpo de esta tabla se dan las probabilidades de una cola conforme a H_0 para la prueba binomial cuando $P = Q = 1/2$. Para ahorrar espacio se omitieron los puntos decimales en las p.

N \ x	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
5	031	188	500	812	969	†										
6	016	109	344	656	891	984	†									
7	008	062	227	500	773	938	992	†								
8	004	035	145	363	637	855	965	996	†							
9	002	020	090	254	500	746	910	980	998	†						
10	001	011	055	172	377	623	828	945	989	999	†					
11		006	033	113	274	500	726	887	967	994	†	†				
12		003	019	073	194	387	613	806	927	981	997	†	†			
13		002	011	046	133	291	500	709	867	954	989	998	†	†		
14		001	006	029	090	212	395	605	788	910	971	994	999	†	†	
15			004	018	059	151	304	500	696	849	941	982	996	†	†	†
16			002	011	038	105	227	402	598	773	895	962	989	998	†	†
17			001	006	025	072	166	315	500	685	834	928	975	994	999	†
18			001	004	015	048	119	240	407	593	760	881	952	985	996	999
19				002	010	032	084	180	324	500	676	820	916	968	990	998
20				001	006	021	058	132	252	412	588	748	868	942	979	994
21				001	004	013	039	095	192	332	500	668	808	905	961	987
22					002	008	026	067	143	262	416	584	738	857	933	974
23					001	005	017	047	105	202	339	500	661	798	895	953
24					001	003	011	032	076	154	271	419	581	729	846	924
25						002	007	022	054	115	212	345	500	655	788	885

29. INTRODUCCION AL ANALISIS MULTIVARIABLE

NOCION

Las técnicas de análisis estadístico, estudiadas hasta aquí, tienen como característica común haberse referido a una sola o dos variables relacionadas entre sí.

Después de considerar el análisis uni- y bivariable que estudian, respectivamente, la variación de una variable o de dos conjuntamente, queda por examinar el análisis multivariable, que pretende describir y explicar las relaciones en su variación de más de dos variables.

Este análisis comprende, como señala Jadish N. Sieth (1980, 131) "el conjunto de todos los métodos estadísticos que en forma simultánea analizan más de dos variables en una muestra de observaciones". Estos métodos constituyen, como señala dicho autor, una ampliación de los análisis uni- y bi-variable.

La realidad social, como ya se ha indicado, es muy compleja; complejidad que se manifiesta en la multiplicidad de variables que generalmente intervienen en los fenómenos sociales. Por ello, el análisis multivariable que, no está restringido de entrada a un número determinado de variables como el uni- o bivariable, puede permitir en principio el estudio de la realidad social de una manera más completa y adecuada.

Característica típica del análisis multivariable es que su complejidad matemática aumenta progresivamente con el número de variables que intervienen en él. Esta complejidad matemática hacía antes inviable la extensión de su práctica habitual. Precisamente, ha sido la difusión y perfección creciente de los ordenadores, lo que está haciendo posible, en la actualidad, dicha extensión. Normalmente, pues, la aplicación de estas técnicas depende y está condicionada por la preparación o existencia previa de los programas pertinentes de ordenadores para llevarlas a cabo.

Es preciso llamar la atención sobre la relatividad de estas técnicas estadísticas y de los modelos matemáticos en la investigación social. Con ello, se pretende prevenir contra la posible creencia de que la simple aplicación de los mismos garantice la exactitud e importancia de los resultados obtenidos.

En primer lugar, se ha de tener en cuenta que los procedimientos estadísticos y matemáticos tienen, generalmente, sus coeficientes de error particulares, que a veces, sobre todo en los modelos matemáticos, pueden ser bastante elevados.

En cuanto a la importancia de los resultados, ésta no se deriva de la simple aplicación de las técnicas estadísticas y matemáticas, sino que depende y está condicionada por el grado en que cumplan los datos utilizados los presupuestos exigidos para la aplicación correcta de cada técnica, por la calidad y exactitud de dichos datos, por la clasificación y tabulación adecuada de los mismos y por la correspondencia de las relaciones estadísticas y matemáticas establecidas con relaciones significativas e importantes en la realidad, de las variables estudiadas.

Por ello, como por lo demás ocurre en todos los campos, la utilidad y eficacia de estas técnicas depende, en último término, de la propiedad, inteligencia, imaginación y sentido crítico con que se las aplique.

CLASIFICACION

En las técnicas de análisis multivariable se tiende a distinguir actualmente dos grandes grupos:

El primero está formado por las técnicas más antiguas en el tiempo de carácter confirmatorio o decisorio, y el segundo por las más modernas de carácter exploratorio e inventivo, y que suele recibir el nombre de análisis de datos.

Así, por ejemplo, Ph. Cibois (1981, 344) escribe que "en sociología hay métodos "durkheimnianos" y métodos "weberianos, los primeros ligados a los coeficientes de correlación y a otros métodos que los utilizan (análisis de regresión y *path analysis*), y los segundos ligados a métodos gráficos que presentan modalidades de configuraciones de respuestas –datos– *(multidimensional scaling)* y análisis de correspondencias".

Las técnicas primeras, presentan una relación más estrecha, como apunta Cibois, con las técnicas estadísticas tradicionales. Por lo general, pretenden estudiar las relaciones –frecuentemente de influencia– entre diversas variables independientes con una, o a veces más de una, dependiente.

Las segundas han surgido después de las primeras, en cierto modo, como reacción ante el cansancio derivado de la aplicación de éstas. Su objetivo no es establecer de manera precisa la determinación de unas variables por otras, sino clasificar científicamente y representar gráficamente el conjunto de datos obtenido en una investigación, con el fin de interpretarlos y descubrir sus propiedades, características y posibles relaciones.

De ahí el carácter heurístico, de invención o hallazgo de idea o sugerencias, y a posteriori, puesto de manifiesto por Benzecri (1980, 12), que tienen estas nuevas técnicas, frente al carácter apriorista, comprobatorio y decisionista con relación a hipótesis determinadas, que ofrecen las más antiguas. De ello se deriva la relación de aquéllas, conforme señala Sonquist (1975, 1), con la fase inductiva del desarrollo del modelo más que con la deductiva de su prueba, peculiar en las tradicionales.

Las técnicas que he llamado clásicas son, sin duda, más rigurosas y concluyentes matemáticamente que las modernas, pero por ello mismo exigen también generalmente, mayor precisión en el nivel de medida de las variables. Dado que en las ciencias sociales, los niveles de medición de las variables propiamente sociales están todavía lejos de ser precisos, se puede aplicar a ellas, en cierto modo, la afirmación que hacen Coole y Lohnes (1971,5) que consideran en las ciencias jóvenes los usos heurísticos, inventivos, de las técnicas de análisis multivariables, más importantes que sus aplicaciones en el contraste de hipótesis.

En sentido similar, A. Fielding (1979, v. 1, 221) opina que en las encuestas el enfoque del análisis de datos debe ser a menudo exploratorio más bien que inferencial. Lo que se requiere son técnicas que resuman adecuadamente y describan las dependencias estructurales en los datos y que se orienten a descubrir regularidades empíricas. Estas constituyen un preludio de la generación de teorías y de la construcción de modelos.

De todas las maneras, no se deben contraponer ambos tipos de técnicas, de tal modo que se consideren incompatibles. Antes al contrario, de acuerdo con Tukey (1977, VII), uno de los pioneros con su análisis exploratorio de esta orientación, si en un principio las técnicas estadísticas fueron más bien de tipo exploratorio y si después adquieren relieve especial las técnicas estadísticas de prueba de hipótesis, hasta el punto de que las técnicas estadísticas iniciales pasaron a un segundo plano como meramente descriptivas, "hoy, las técnicas exploratorias y confirmatorias pueden, y deberían, avanzar unas al lado de otras".

Igualmente, Fielding subraya la concordancia entre la fase exploratoria y la inferencial. "La exploración de datos, afirma, sugiere hipó-

tesis. A su vez, las hipótesis formales una vez contrastadas pueden sugerir nuevas formas de examinar los datos".

Lo que procede, pues (supuesto que no hay técnicas definitivas y válidas para todo y en todos los casos, y que todas tienen sus límites e imperfecciones, especialmente en las ciencias sociales) es realizar una aplicación inteligente de ellas en los casos para los que sean apropiadas, procurando, siempre que sea posible, sacar partido de lo que tienen de complementarias.

No se debe olvidar, por último, que en la investigación científica las técnicas matemáticas y estadísticas en general no son un fin en sí mismas, sino un instrumento cuya eficacia depende de su apropiada utilización. Como procedimientos formales que son, admiten toda clase de datos, ciertos y erróneos, y sus resultados, incluso correctos operativamente, frecuentemente lo son solamente con un grado de probabilidad mayor o menor, por lo que deben ser siempre objeto de una interpretación racional y crítica, teniendo en cuenta la realidad, la experiencia y la doctrina.

Expuesto lo anterior, las técnicas de análisis multivariable más extendidas en la actualidad se expresan en el siguiente esquema:

TECNICAS DE ANALISIS MULTIVARIABLE

- **Confirmatorios** Análisis (de)
  - Componentes principales
  - Regresión múltiple
  - Correlaciones canónicas
  - Discriminante
  - Factorial
  - Varianza
  - Covarianza
  - Senderos (path)
  - Log-lineal
  - Sistemas
- **Exploratorios** Análisis (de)
  - Clusters
  - Correspondencias
  - Estructura latente
  - Escalonamiento multidimensional

Todas estas técnicas, según se ha indicado, son bastante complicadas y laboriosas operativamente, y tanto más cuanto mayor número de variables y observaciones sobre cada variable comprendan. Por ello, salvo en casos muy simples, su aplicación sólo resulta viable mediante la utilización de ordenadores para realizarlas, utilizando los programas ya existentes.

De acuerdo con lo anterior, a continuación se hace mención únicamente a la noción y aplicaciones básicas de cada uno de estos tipos de análisis, pues, como señala K. Hope (1972, 9) el investigador debe conocer sus aplicaciones y comprender su significado. No obstante, el análisis de senderos serán objeto de consideración especial al tratar el análisis causal.

ANALISIS DE COMPONENTES PRINCIPALES

Se orienta al análisis de las matrices multivariables de varianzas y covarianzas con el fin de descubrir sus componentes principales, es decir, las variables implícitas, que tienen un mayor poder explicativo en la variación total del sistema. Su objetivo es, pues, determinar estos componentes y hacer factible, de este modo, una posible reducción de las variables del sistema a estos componentes, omitiendo las demás cuando se pueda prescindir de ellas con escasa pérdida de información.

Se trata de una técnica básica que tiene aplicación en otros tipos de análisis, como, por ejemplo, el análisis factorial y el de regresión. Sus aplicaciones principales consisten en:

a) Reducir un conjunto de datos a la síntesis de unos pocos ítems destacados, y

b) Mostrar qué variables pueden ser omitidas en un sistema sin alterar demasiado la información básica.

ANALISIS DE REGRESION MULTIPLE

Consiste en la extensión del análisis de regresión a más de dos variables, una dependiente y las demás independientes. Pretende especificar el grado en que cada una de las distintas variables independientes determina la variación de la variable dependiente en cuestión.

ANALISIS DE CORRELACIONES CANONICAS

Es una extensión del análisis de regresión múltiple al caso especial en que exista, no una, sino varias variables dependientes y busca, como el análisis factorial, establecer la relación estadística entre un conjunto de variables independientes y otro de variables dependientes, mediante la determinación de las correlaciones máximas de dichas variables con las variables implícitas que subyacen a ambas.

ANALISIS FACTORIAL

De acuerdo con Blalock (1977, p. 417) como fundamento del análisis factorial se encuentra "la idea de que si tenemos un gran número de índices o variables correlacionadas entre sí, estas relaciones recíprocas pueden deberse a la presencia de una o más variables o factores subyacentes relacionados en grado diverso con aquéllos".

Se supone que las intercorrelaciones altas dentro de un grupo de variables se deben a uno o varios factores o variables generales.

Precisamente uno de los objetivos principales del análisis factorial es identificar estos factores o variables comunes y por tanto más generales que los dados. Con ello frecuentemente reciben significación muchos conjuntos de correlaciones que, de otra manera, parecía que carecían de sentido.

El análisis factorial está sobre todo justificado, cuando, como dice Schuessler (1971, p. 44), nos encontramos con un conjunto de intercorrelaciones significativas entre variables sociales y económicas muy distintas, por ejemplo, riqueza, empleo, desarrollo, divorcio, criminalidad, escolarización, etc., con referencia a ciudades o núcleos de población, que dan la impresión de constituir una mezcla de cifras sin un sentido coherente.

Entonces es oportuno utilizar el análisis factorial para determinar si dicho conjunto de intercorrelaciones aparentemente inconexas se deben a uno o varios factores o variables no explícitas con los cuales las variables iniciales se hallan fuertemente correlacionadas.

ANALISIS DISCRIMINANTE

Tiene por objeto establecer, dado un conjunto de variables, los coeficientes que discriminan de modo óptimo, con relación a dichas variables, a dos sectores de la población investigada, por ejemplo, los que no van y van al teatro, los votantes y no votantes del partido X, los que tienen y no tienen TV. El análisis discriminante es aplicable, por tanto, para determinar las propiedades o mejor dicho pesos que caracterizan a los dos subgrupos de que se trate respecto a las variables tenidas en cuenta. Asimismo, estos pesos o coeficientes pueden servir para clasificar a individuos, que no se sabe a qué subgrupo pertenecen, cuando se conocen sus medidas en las variables del análisis. Como señalan Zaltman y Burger (1980, v. 2, 117) el MDA, análisis discriminante múltiple, tiene un uso cada vez mayor en marketing, donde se emplea para descubrir diferencias entre territorios de venta, para saber qué clase de personas tienen el coche X, etc.

ANALISIS DE VARIANZA MULTIPLE

Es una extensión del análisis de varianza simple, cuando existe, no una, sino dos o más variables independientes. Consiste, como tal, en la comparación de las varianzas de las diferentes categorías de datos de las variables del análisis, con el fin de determinar si existen o no diferencias significativas en los valores medios de las distintas categorías de datos.

Tuvo su origen en las investigaciones agrícolas. Con él se trataba de decidir si los distintos tratamientos aplicados a parcelas (diferencias de abonos, semillas o cultivos) daban lugar o no a diferencias de rendimiento significativas.

ANALISIS DE COVARIANZA

Este análisis combina el análisis de varianza y el de regresión. Consiste en realizar, partiendo de una o más variables nominales y dos o más variables de intervalo, el análisis de varianza entre las variables nominales (independientes) y una variable de intervalo, controlando o manteniendo constante la restante o restantes variables de intervalo.

ANALISIS DE SENDEROS

Consiste en la repetición sucesiva del análisis de regresión múltiple respecto a cada una de las variables del modelo. Tiene aplicación en el análisis causal, porque permite estimar los parámetros de los modelos causales, que se interpretan como indicadores de la cantidad de cambio (estandarizado) en una variable dependiente que es atribuible al cambio (estandarizado) de una variable anterior.

ANALISIS DE SISTEMAS

Su objetivo es el estudio del cambio de un sistema derivado de su funcionamiento. El funcionamiento del sistema se realiza por los elementos del sistema, que hallándose en determinado estado y recibiendo ciertos impulsos o entradas, efectúan operaciones que dan lugar a salidas o productos del sistema y a nuevos estados del mismo. Tiene aplica-

ción en el estudio de toda clase de sistemas de cualquier tipo: industriales, administrativos, sociales, etc.

ANALISIS LOG-LINEAL

Es una nueva técnica específica, válida para el análisis unitario y global de las relaciones de influencia o de la estructura causal de las variables que integran una tabla de contingencia determinada. De acuerdo con C. Payne (1979, 104) el objetivo que se pretende con este análisis es el contraste de hipótesis sobre las relaciones entre las variables, mediante la búsqueda y discernimiento del modelo que mejor se adapte a los datos de una tabla.

Sus aplicaciones se extienden a todas las tablas de contingencia y no sólo al estudio matemático de modelos causales, sino también al de las diversas tablas de rotación, que representan cambios a lo largo del tiempo.

ANALISIS DE CLUSTER

Es una técnica para realizar clasificaciones, como se indica en su capítulo, pero que también se puede utilizar para determinar y analizar los distintos componentes de una población.

ANALISIS DE CORRESPONDENCIAS

Se trata de una técnica aplicable al análisis de matrices de datos y, en general, de tablas de contingencia. Ideado por J.P. Benzécri, se orienta a descubrir la forma y modelo de datos multivariables y de sus relaciones. Su objetivo es situar las distintas categorías de datos analizados en un plano cartesiano y estudiar cómo las tablas de datos se organizan alrededor de sus centros de gravedad correspondientes. Por ello, da lugar a representaciones gráficas donde los datos se convierten en puntos sobre un eje o en un plano.

ESCALONAMIENTO MULTIDIMENSIONAL

Se basa en la determinación de distancias entre datos y su objetivo es obtener un mapa que represente espacialmente la estructura de los datos, como si, dadas las distancias entre las distintas localidades, pretendiéramos construir a partir de ellas el mapa que exprese su respectiva

situación geográfica. Se puede aplicar en toda clase de datos, si bien su utilización más natural tiene lugar en los datos que expresan de por sí un juicio o evaluación de objetos diversos.

EL ANALISIS MULTIVARIABLE Y LOS MODELOS MATEMATICOS

Esta relación se puede concretar diciendo, como lo hace Sonquist, que los análisis multivariables de tipo exploratorio tienen conexión con la fase inductiva de la búsqueda y desarrollo de modelos y los de tipo confirmatorio con la deductiva de su prueba. Es, por tanto, la realización de los tipos de análisis multivariable confirmatorio la que supone la aplicación y desarrollo de los modelos matemáticos (Cfr. su noción en el cap. 3). De ahí que sea oportuno exponer aquí las operaciones básicas que comprende tal aplicación y resolución.

OPERACIONES QUE COMPRENDEN LOS MODELOS

En el desarrollo completo de un modelo matemático se pueden distinguir las siguientes operaciones:

1. Formulación teórica del modelo.
2. Representación gráfica.
3. Especificación de las ecuaciones matemáticas que se cree corresponden a las relaciones entre las variables en el modelo teórico.
5. Estimación de los parámetros de sus ecuaciones.
6. Verificación, prueba, o test del modelo, y
7. Evaluación del modelo.

1. *Formulación teórica del modelo.*— Es la representación teórica de un fenómeno o sector de la realidad con la especificación de las variables, y relaciones principales entre ellas, que se cree configuran dicho fenómeno o sector de la realidad.

2. *Representación gráfica.*— El modelo se representa gráficamente haciendo figurar, generalmente, sus elementos por puntos, y las relaciones entre estos elementos por flechas.

3. *Especificación del modelo.*— Esta operación, que los norteamericanos suelen llamar *mapping* tiene por objeto, como señalan Leik y Meeker (1975, p. 11) establecer las relaciones de correspondencia entre las matemáticas por un lado y la teoría y los datos por otro.

Requiere que se especifique matemáticamente, mediante las correspondientes ecuaciones, la naturaleza y la forma matemática de las relaciones entre las variables del modelo.

Esta representación sólo se puede realizar en términos aproximados y supone un compromiso entre la necesidad de simplificación en los modelos y una correspondencia precisa en lo posible entre las relaciones matemáticas y las relaciones reales.

4. y 5. *Obtención de datos empíricos sobre las variables del modelo y estimación de los parámetros de las ecuaciones.*— Especificadas las ecuaciones del modelo se precisa seguidamente obtener datos estadísticos o de encuesta sobre la dimensión empírica de las variables del modelo.

Estos datos deben servir para la estimación del valor de los parámetros de las ecuaciones, con lo que el modelo quedará totalmente especificado.

6. *Verificación, prueba o test del modelo.*— Se pueden distinguir dos tipos de prueba del modelo principales.

Primero, mediante la comparación de los valores estimados de la variable dependiente que resulten según los parámetros hallados de las ecuaciones, con los valores de la misma variable dependiente obtenidos empíricamente. Esta comparación se puede realizar, por ejemplo, hallando el coeficiente de determinación r^2 entre las dos series de valores.

Segunda, en el caso de que la prueba anterior sea satisfactoria, se puede aplicar el modelo a datos obtenidos de otras poblaciones para comprobar si sigue siendo válido el modelo para ellas.

Asimismo, se pueden ensayar otros tipos de ecuaciones o relaciones entre los datos del modelo, para ver cuál de ellas se ajusta mejor a los datos empíricos.

7. *Evaluación del modelo.*— La evaluación del modelo se puede efectuar:

a) determinando su consistencia matemática, es decir, que las ecuaciones matemáticas del modelo son firmes y correctas desde este punto de vista;

b) su adecuación empírica, o sea, el ajuste entre los resultados derivados del modelo y los resultados obtenidos empíricamente para las variables investigadas; y

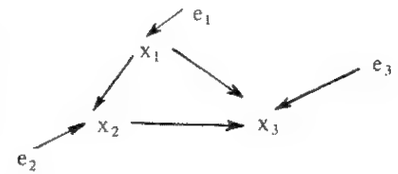
c) su valor teorético. Lo tendrá, si, como señalan Leik y Meeker (1975, p. 33) nos dice algo nuevo, si explica mejor el fenómeno que los

modelos disponibles antes, si propociona la forma de incorporar teorías previamente distintas y si proporciona soluciones teoréticas de interés para los sociólogos.

EJEMPLO DE CONSTRUCCION DE UN MODELO SOCIOLOGICO

Según la investigación social de Durkheim sobre el suicidio, dos de las variables que influyen en él son la edad y el estado civil. A su vez es claro que la edad influye en el estado civil, y que el suicidio no influye en la edad ni en el estado civil, ni éste en la edad.

Por tanto se puede representar esta teoría gráficamente con el siguiente modelo gráfico, en el que las flechas expresan las relaciones de influencia indicadas entre las variables.



Matemáticamente se puede representar este modelo gráfico por el siguiente sistema de ecuaciones.

Vemos que en x_1 no influye ninguna otra variable. Sin embargo pueden existir otras variables implícitas, no conocidas o no tenidas en cuenta en el modelo, que influyen en esta variable. Esta influencia se suele llamar error y se representa por e_1. Por tanto la ecuación de esta variable sería la siguiente:

$$x_1 = e_1$$

En x_2 influye según el gráfico x_1, y además hay que tener en cuenta, como antes, la posible influencia de otras variables desconocidas. Luego la ecuación será en este caso:

$$x_2 = a_{12}\ x_1 + e_2$$

donde a_{12} es el coeficiente o parámetro que representa cuantitativamente el grado de influencia de x_1 en x_2.

Por último en la variable x_3, el suicidio, influyen a la vez, en un grado determinado expresado por los parámetros o coeficientes a_{ij}, las variables x_1 y x_2 y el error propio de esta variable. La ecuación pues es:

$$x_3 = a_{13}\ x_1 + a_{23}\ x_2 + e_3$$

Con esta última ecuación ya tendríamos especificado matemática y totalmente el modelo teórico. Para resolverlo, necesitaríamos únicamente obtener datos empíricos sobre las variables del modelo y estimar

de acuerdo con los valores obtenidos de las mismas los parámetros de las ecuaciones.

Adviértase que en las ecuaciones de este modelo, que es uno de los más simples que pueden darse, se supone en todos los casos que la relación que une a unas variables con otras es lineal.

Como ejercicios de este capítulo se ofrecen a continuación algunos problemas sobre la formación de las matrices de,

- Sumas de cuadrados y productos,
- De varianzas y covarianzas, y
- De correlaciones

Así, como sobre la suma, resta, multiplicación, inversión de matrices y hallazgo de valores propios, procedimientos de cálculo que son básicos en muchas de las formas de análisis multivariable.

EJERCICIOS

226

Dados los siguientes valores de una muestra de cinco individuos en tres variables: x_1, x_2, x_3

	x_1	x_2	x_3
A	2	9	14
B	14	20	8
C	8	17	11
D	16	11	10
E	10	13	2
Σ	50	70	45
$\bar{X}$	10	14	9

Hallar la matriz de sumas de cuadrados y productos y para ello,

a) Normalizar los valores de dichas variables, expresándolos como desviaciones de su media.

RESPUESTA

Hay que sustituir simplemente los valores iniciales en cada variable por la diferencia de los mismos con su media respectiva.

Valores originales			*Valores normalizados*		
x_1	x_2	x_3	x_1	x_2	x_3
2	9	14	−8	−5	5
14	20	8	4	6	−1
8	17	11	−2	3	2
16	11	10	6	−3	1
10	13	2	0	−1	−7
50	70	45			
10	14	9			

b) Hallar las sumas de cuadrados y productos de los valores de cada variable normalizados.

RESPUESTA

Se elevan al cuadrado cada uno de dichos valores y se suman los de la misma variable. Asimismo se multiplican los valores de cada una de las variables sucesivamente por los de las demás y se suman como antes.

	x_1	x_1^2	x_2	x_2^2	x_3	x_3^2	$x_1 x_2$	$x_1 x_3$	$x_2 x_3$
	−8	64	−5	25	5	25	40	−40	−25
	4	16	6	36	−1	1	24	−4	−6
	−2	4	3	9	2	4	−6	−4	6
	6	36	−3	9	1	1	−18	6	−3
	0	0	−1	1	−7	49	0	0	−7
Σ		120		80		80	40	−42	−35

c) Formar la matriz de sumas y productos.

RESPUESTA

Se trata de una matriz cuadrada formada por las variables de que se trata, x_1, x_2, x_3 como cabeceras de sus filas y columnas. Por tanto, en la casilla de la diagonal

principal se tendrá las sumas de los cuadrados o multiplicación de cada variable por sí misma y en las restantes casillas las sumas de los productos de los valores de las variables que se crucen en cada casilla:

Sumas: $x_1^2 = 120$; $x_2^2 = 80$; $x_3^2 = 80$; $x_1\ x_2 = 40$; $x_1\ x_3 = -42$; $x_2\ x_3 = -35$

	x_1	x_2	x_3
x_1	120	40	-42
x_2	40	80	-35
x_3	-42	-35	80

227

Dada la matriz de sumas y cuadrados de productos calculada en el ejercicio anterior, hallar la matriz de varianzas y covarianzas.

RESPUESTA

La matriz pedida se obtiene simplemente de la misma matriz de sumas y productos del ejercicio anterior, dividiendo cada uno de los valores de sus casillas por N-1, en este caso 5-1 = 4.

	x_1	x_2	x_3
x_1	30	10	10,5
x_2	10	20	8,75
x_3	10,5	8,75	20

Los valores de la diagonal principal son las varianzas y los restantes las covarianzas.

228

Dada la matriz de varianzas y covarianzas determinada en el ejercicio anterior, calcular la matriz de correlaciones.

RESPUESTA

En esta matriz los valores de la diagonal principal indican la correlación de cada variable consigo mismo que, por principio, es total o igual a la unidad. La correlación de las distintas variables entre sí se obtiene dividiendo cada covarianza de la matriz de varianzas y covarianzas por el producto de las raíces cuadradas de las varianzas que corresponden, en la diagonal de la matriz de varianzas y covarianzas, a la fila y la columna perteneciente a cada covarianza.

Se tendrán, por tanto, las siguientes correlaciones:

$$r_{12} = r_{21} = \frac{10}{\sqrt{30}\ \sqrt{20}} = .41$$

$$r_{13} = r_{31} = \frac{10,5}{\sqrt{30}\ \sqrt{20}} = \frac{-10.5}{24.51} = -.43$$

$$r_{23} = r_{32} = \frac{-8,75}{\sqrt{20}\ \sqrt{20}} = \frac{-8,75}{20.07} = -.43$$

	x_1	x_2	x_3
x_1	1	41	-.43
x_2	.41	1	-.43
x_3	-.43	-.43	1

229

Dadas las siguientes matrices, hallar su suma y su resta.

$$\mathbf{A} = \begin{pmatrix} 4 & -3 \\ 5 & 1 \\ 6 & 2 \end{pmatrix} \quad \text{y} \quad \mathbf{B} = \begin{pmatrix} 12 & 6 \\ -8 & -3 \\ 9 & 0 \end{pmatrix}$$

RESPUESTA

Una matriz es un conjunto de filas y columnas de números, con la que se opera considerándola como una unidad de cálculo.

$$\mathbf{A} + \mathbf{B} = \begin{pmatrix} 16 & 3 \\ -3 & -2 \\ 15 & 2 \end{pmatrix}$$

$$\mathbf{A} - \mathbf{B} = \begin{pmatrix} -8 & -9 \\ 13 & 4 \\ -3 & 2 \end{pmatrix}$$

Las matrices se pueden sumar si tienen el mismo número de filas y de columnas. La matriz suma está formada por las sumas algebráicas de los elementos correspondientes de las matrices sumadas, y la matriz diferencia por la resta.

Las matrices en su conjunto se designan por letras mayúsculas en negrita.

230

Dada la matriz siguiente, escribir su traspuesta.

$$\begin{pmatrix} 4 & -3 \\ 5 & 1 \\ 6 & 2 \end{pmatrix}$$

RESPUESTA

La traspuesta **A'** de una matriz **A** es la nueva matriz que se obtiene intercambiando en **A** las filas por las columnas.

$$\begin{pmatrix} 4 & 5 & 6 \\ -3 & 1 & 2 \end{pmatrix}$$

231

Dadas las siguientes matrices, premultiplicarlas primero y postmultiplicarlas después.

$$\mathbf{A} = \begin{pmatrix} 4 & -3 \\ 5 & 1 \\ 6 & 2 \end{pmatrix} \quad y \quad \mathbf{B} = \begin{pmatrix} -8 & 13 & 3 \\ -9 & 4 & 2 \end{pmatrix}$$

RESPUESTA

La multiplicación de matrices no tiene la propiedad transitiva. Es decir, en ella no es lo mismo la multiplicación de A por B que la de B por A. Por ello, se distingue entre la premultiplicación de A por B, que da como resultado BA, y postmultiplicación de B por A que da como resultado AB.

La multiplicación de matrices exige que una matriz tenga las mismas filas que columnas la otra y se realiza multiplicando los elementos de cada fila por los elementos de cada columna sucesivamente y sumando los productos que constituyen los elementos de la nueva matriz multiplicada.

AB =

$4(-8) + (-3)\,-9 = -5$; $5(-8) + 1(-9) = -49$

$6(-8) + 2(-9) = -66$; $4(13) + -3(4) = 30$;

$5(13) + 1(4) = 69$; $6(13) + 4(2) = 86$;

$4(3) + (-3)\,2 = 6$; $5(3) + 1(2) = 17$; $6(3) + 2(2) = 13$

$$\mathbf{AB} = \begin{pmatrix} -5 & 30 & 6 \\ -49 & 69 & 17 \\ -66 & 86 & 13 \end{pmatrix}$$

$\mathbf{BA} = -8(4) + 13(5) + 3(6) = 51$

y así, sucesivamente.

$$\mathbf{BA} = \begin{pmatrix} 51 & 43 \\ -4 & 33 \end{pmatrix}$$

232

Dada la siguiente matriz, hallar su matriz escalar, tomando como factor escalar 5.

$$\begin{pmatrix} 51 & 43 \\ -4 & 33 \end{pmatrix}$$

RESPUESTA

Matriz escalar es la que resulta si se multiplica cada uno de los elementos de la matriz inicial por el factor dado.

$$5 \begin{pmatrix} 51 & 43 \\ -4 & 33 \end{pmatrix} = \begin{pmatrix} 255 & 215 \\ -20 & 165 \end{pmatrix}$$

233

Dada la siguiente matriz, hallar su determinante.

$$\begin{pmatrix} 4 & 7 & 3 \\ 2 & 0 & 4 \\ 1 & 4 & 6 \end{pmatrix}$$

RESPUESTA

Las matrices cuadradas, con igual número de filas que de columnas, tienen un determinante, que se simboliza por $|\mathbf{A}|$ si la matriz es A.

El determinante de una matriz es un valor singular que determina la matriz.

Para calcular su valor, se multiplican entre sí sucesivamente los elementos de la matriz, de modo que en cada multiplicación se elija un solo elemento de cada fila y columna. La suma algebraica de los productos resultantes es el determinante. El signo de los productos sumados es más si, colocados los elementos según el orden de sus filas, el número de inversiones de este orden de colocación respecto al orden natural de las columnas es par, y menos si es impar.

$$|\mathbf{A}| = (4.0.6) - (4.4.4) - (7.2.6) + (7.4.1) + (3.2.4) - (3.0.1) = 24 - 64 - 84 + 28 + 24 - 3 = -75$$

Vemos que en cada producto sus elementos pertenecen todos a distintas columnas y filas. En cuanto al signo, por ejemplo, en 7.2.6, el 7 de la columna 2, precede una vez al 2 de la columna 1, luego se altera el orden natural y hay una inversión y el signo es negativo. Igualmente en 3.2.4 el 3 de la columna 3 precede al 2 y al 4 de las columnas 1 y 2, luego hay dos inversiones, y el signo es positivo.

234

Dada la siguiente matriz, hallar su matriz inversa.

$$\mathbf{A}\begin{pmatrix} 3 & 4 & 1 \\ 2 & 6 & 3 \\ 8 & 1 & 2 \end{pmatrix}$$

RESPUESTA

La inversa de la matriz cuadrada **A** es una matriz cuadrada $\mathbf{A}^{-1}$, tal que multiplicada **A** por $\mathbf{A}^{-1}$, da como resultado la matriz de identidad, o matriz unidad, I, formada por unos en la diagonal principal y ceros en todas las demás casillas.

Para hallarla, primero se calculan los cofactores de la matriz que son matrices de orden inferior obtenidas cuando se prescinde en la matriz sucesivamente de una fila determinada y la columna correspondiente.

Con los cofactores de la primera fila se puede hallar primero el determinante de la matriz de la forma siguiente:

$$|A| = 3\begin{vmatrix} 6 & 3 \\ 1 & 2 \end{vmatrix} - 4\begin{vmatrix} 2 & 3 \\ 8 & 2 \end{vmatrix} + 1\begin{vmatrix} 2 & 6 \\ 8 & 1 \end{vmatrix} = 61$$

Después se expanden todos los cofactores de la matriz, es decir, de todas las filas que serían los siguientes:

$$\mathbf{A} = \begin{matrix} \begin{vmatrix} 6 & 3 \\ 1 & 2 \end{vmatrix} - \begin{vmatrix} 2 & 3 \\ 8 & 2 \end{vmatrix} + \begin{vmatrix} 2 & 6 \\ 8 & 1 \end{vmatrix} + \begin{vmatrix} 2 & 6 \\ 8 & 1 \end{vmatrix} \\ \begin{vmatrix} 4 & 1 \\ 1 & 2 \end{vmatrix} + \begin{vmatrix} 3 & 1 \\ 8 & 2 \end{vmatrix} - \begin{vmatrix} 3 & 1 \\ 8 & 4 \end{vmatrix} - \begin{vmatrix} 3 & 4 \\ 8 & 1 \end{vmatrix} \\ \begin{vmatrix} 4 & 1 \\ 6 & 3 \end{vmatrix} - \begin{vmatrix} 3 & 1 \\ 2 & 3 \end{vmatrix} + \begin{vmatrix} 3 & 4 \\ 2 & 6 \end{vmatrix} + \begin{vmatrix} 3 & 4 \\ 2 & 6 \end{vmatrix} \end{matrix}$$

A continuación se calculan los determinantes de cada cofactor, que aquí serán:

$$\mathbf{A} = \begin{pmatrix} 9 & -20 & -46 \\ -7 & -2 & 29 \\ 6 & -7 & 10 \end{pmatrix}$$

En el caso de la matriz de dos columnas y dos filas, el determinante es igual a la diferencia entre los productos de los elementos de la diagonal principal y los de la secundaria.

Después se determina la traspuesta de la matriz de determinantes obtenida, que será:

$$\mathbf{A}' = \begin{pmatrix} 9 & -7 & 6 \\ -20 & -2 & -7 \\ -46 & 29 & 10 \end{pmatrix}$$

Finalmente, la matriz resultante de dividir cada elemento de esta matriz traspuesta por el determinante, es la matriz inversa buscada.

$$\mathbf{A}^{-1} = \begin{pmatrix} 9/61 & -7/61 & 6/61 \\ -20/61 & -2/61 & -7/61 \\ -46/61 & 29/61 & 10/61 \end{pmatrix}$$

30. ANALISIS CAUSAL

METODOS EXPERIMENTALES Y CORRELACIONALES

Como se ha indicado ya, el análisis causal es el que pretende investigar las relaciones de influencia o causalidad entre variables.

Respecto al estudio de la causalidad, se ha establecido una distinción importante entre los llamados métodos experimentales y los métodos correlacionales.

Los primeros se basan, como es sabido, en la reducción de los complejos sistemas de causalidad reales a sistemas sencillos, si es posible sólo de dos variables, mediante el aislamiento y control experimental de las restantes variables.

Otra cosa ocurre en las investigaciones no experimentales, donde no existe el aislamiento y el control experimental. En este caso, lo normal es la complejidad de las relaciones y la necesidad de acercarse a ellos mediante modelos con varias variables. Estos modelos se suelen basar en el análisis estadístico de las correlaciones entre las variables del modelo, de ahí que, en contraposición a los experimentales, se les da el nombre de métodos correlacionales.

Dada su consideración aludida como modelos causales, implican y se basan de alguna manera en la idea de causalidad, lo que hace conveniente detenerse un poco para examinar el sentido de la relación de causalidad o causa.

RELACION DE CAUSALIDAD

Desde un punto de vista filosófico, ontológico, se puede entender como causa aquello que hace ser a algo que no es o que venga a ser de forma distinta lo que es.

En otras palabras, causa es aquello capaz de producir un nuevo ser o de modificar la manera de ser de lo que ya es.

Esta noción de causa implica una distinción importante, sin duda, en la investigación científica, entre la causa que produce algo nuevo de aquella que sólo modifica lo ya existente.

Habida cuenta de que la investigación científica no se ocupa de los seres o cosas en su conjunto o en la totalidad de su ser, sino sólo de las características, variables, de esos seres o cosas en los estudios descriptivos y de las relaciones entre dichas características o variables en los explicativos, es obvio que a la ciencia le interesa fundamentalmente la causalidad en el segundo sentido, de aquello que produce no un nuevo ser sino la modificación de lo existente en algunas de sus características.

Por tanto, cuando se dice en la ciencia que dos características de las cosas o fenomenos, o lo que es lo mismo dos variables, están unidas por una relación de causalidad, se quiere decir, en último término, que una variable influyen en la otra, en el sentido de que una modificación en la primera produce o da lugar a una modificación en la segunda.

En este sentido, Ph. van Parijs en su trabajo "La syntase de l'explicatión dans les sciences sociales (Recherches Sociologiques, 8, 2, 199, p. 277) dice que hay una relación causal cuando modificando A se puede producir una modificación en B.

En todo caso, bien se trate de la producción de un nuevo ser o sólo de la modificación de sus características, la noción de causalidad implica siempre un componente ontológico, en el sentido de creación o transformación de algo en cuanto a su ser, que hace que la causalidad sea, en sí, algo metafísico, más alla de la física y de los procedimientos de observación actuales de la ciencia.

El campo de la ciencia es el de la realidad observable. Por ello con relación a la causalidad, en la aparición de un nuevo ser o su transformación, científicamente, sólo se puede observar empiricamente que un ser que antes no era es o que antes era de una forma lo que ahora es de otra, y no la transformación ontológica, en el ser, que supone este cambio.

De aquí que, científicamente, las definiciones de causa y relación causal se centren en esta idea de sucesión necesaria, constante e irreversible.

J. A. Davis escribe (1972, p. 108) que "dos variables tienen una relación causal cuando una reordenación de casos de una va seguida de una reordenación de casos en la otra". Y Stefan Novak (1975, p. 82)

afirma que "decir que S es la causa de B significa que donde quiera que y siempre que, dentro de los límites de la generalización causal de que se trate, S ocurre (o debería ocurrir) es (o debería ser) seguido por B, independiente de si S ocurre o (debería ocurrir) espontáneamente o fue (o debería ser) traído a la existencia por alguna acción voluntaria o por cualquier "actor" o "producto".

Una consecuencia importante de esta idea científica de causa, es que la causalidad no se puede comprobar de manera inductiva, empírica, sino que supone una inferencia deductiva.

La causalidad, como señalan Jahoda y otros (1970 4ª ed., p. 471) implica tres tipos de evidencia empírica:

1. Que X e Y varían concomitantemente en la forma descrita por la hipótesis.
2. Que Y (efecto) no precede a X en el tiempo.
3. Que no haya otros factores que determinen Y.

Pero ocurre que nunca se puede estar seguro empíricamente de que X e Y varíen *siempre* concomitantemente en la forma descrita en la hipótesis así como de la no existencia de otros factores que determinen Y. De ahí que la inferencia causal, sobre la existencia de una relación de causalidad entre dos variables, reclame un cuarto elemento de carácter racional y deductivo, que se podría caracterizar genéricamente como la existencia de una teoría según la cual parezca racional y lógica la relación de causalidad en cuestión.

Como dice Parijs (loc. cit, p. 228) "incluso en la experimentación más cuidada, uno no está nunca seguro de que todas las variables susceptibles de influir en la variable efecto han sido controladas".

De ahí la necesidad, en su opinión, en cuanto a la explicación científica de las relaciones causales, de que el hecho a explicar sea no solamente deducible por la mención de una regularidad, sino también inteligible por la intuición de un mecanismo capaz de producirlo, que lo haga inteligible por encuadrarlo en un marco de inteligibilidad legítima.

PROCEDIMIENTOS DE ANALISIS DE LOS MODELOS MULTIECUACIONALES CAUSALES EN LAS CIENCIAS SOCIALES

Como se ha indicado no está al alcance de la ciencia, dentro de sus límites, comprobar empíricamente la causalidad, en cuanto al vínculo ontológico que constituye su esencia.

Sin embargo, si lo está comprobar empíricamente si se dan, hasta el presente, los tres tipos de evidencia indicados y la congruencia con una teoría, como elementos que implican la noción indicada de causalidad.

Igualmente, nada impide que los científicos ideen y construyan modelos que sean representación de los sistemas de causalidad real y que intenten verificarlos empíricamente para ver si concuerdan en un cierto grado con la realidad.

Como escribe Blalock (1972, p. 6) "se admite que el pensamiento causal pertenece completamente al nivel teorético y que las leyes causales no pueden demostrarse empíricamente. Pero esto no significa que no sea conveniente discurrir causalmente y desarrollar modelos cuyas implicaciones sean verificables indirectamente".

De ahí que, de hecho, en la práctica, existen en las ciencias sociales diversos procedimientos para el análisis causal. De entre ellos, aquí se hace una referencia básica a los dos siguientes:

1. Análisis multivariable de Lazarfeld, también llamado estrategia de elaboración.
2. Análisis de senderos (*path analysis*)

ANALISIS MULTIVARIABLE DE LAZARSFELD

Ideado por este autor, consiste en la prueba de las relaciones observadas entre dos variables mediante la puesta en relación de ambas con una tercera variable que recibe el nombre de test o variable prueba.

Este análisis se puede decir que se basa en último término en la técnica del experimento. Como en éste, según es sabido, se opera manteniendo constantes o controlando todas las variables o factores que pueden influir en un fenómeno, pues es el único modo de poder determinar que el efecto producido en la variable dependiente en estudio es causado por el estímulo introducido, o variable independiente, así en el análisis multivariable se va probando la relación entre las dos variables en estudio, igualando, haciendo constante o controlando, no ya todas, lo que no es factible en las Ciencias Sociales, pero sí una por una de las demás variables principales que se cree pueden influir en el fenómeno, las cuales se mantienen constantes o se controlan respecto a la variable independiente de la relación original.

Por ejemplo, se puede haber relacionado, genéricamente, la práctica religiosa, alta y baja, con la edad (jóvenes y viejos) pero es probable que en esta relación influya también el sexo. Para probar la anterior relación

se introduce la nueva variable sexo, y se mantiene constante respecto a la variable independiente edad. Con esto se quiere decir que la tabla inicial, en la que no se ha controlado el sexo, muestra la relación entre práctica religiosa y edad, jóvenes-viejos, pero sin separación de sexos, comprendiendo en jóvenes y viejos, sin distinción, a los hombres y las mujeres. Pero la introducción de la nueva variable supone la superación de esta indiscriminación, pues en ella la tabla inicial se desarrolla o desdobla en dos, que muestran la relación de edad y práctica religiosa por un lado en las mujeres sólo y por otro en los hombres. Es decir, manteniendo distinta la edad, jóvenes y viejos, se ha igualado o mantenido constante el sexo en cada caso. A esta operación se le da el nombre técnico de estratificación de la relación original.

Los elementos que intervienen en el análisis multivariable son las variables y las tablas. Las variables son tres, la dependiente y la independiente de la relación original y la variable prueba. Esta puede ser antecedente, o anterior en el tiempo a X e Y, o interviniente si se sitúa en el tiempo entre X variable independiente e Y variable dependiente.

Las tablas son, fundamentalmente, la original y las otras dos llamadas parciales, derivadas de introducir, manteniéndola constante, la variable prueba. También se pueden formar tablas con los totales de los valores absolutos o marginales de las dos tablas parciales, bien correspondientes al pie o a los laterales. Estas tablas relacionan cada una de las variables de la relación original con la variable test.

Todas las tablas indicadas reflejan los resultados del análisis multivariable. Puede suceder o bien que en los parciales (los dos o uno sólo) aparezca una relación más significativa entre X e Y, las variables iniciales, o bien que no aparezca esta relación, y sí en las tablas marginales, poniendo de manifiesto una relación entre la variable prueba T y una de las variables X e Y.

Cruzando estas dos clases de resultados con el carácter de la variable test indicado, antecedente o interviniente, se obtiene la tipología de las especies de análisis multivariables siguiente:

En uno de los parciales aparece una asociación significativa (o más significativa) entre X e Y.

La asociación aparece significativamente entre una de las variables originales (X o Y) y la tercera variable (T) o en el marginal, pero no en los parciales.

		P	*M*
T es antecedente	A	PA = Especificación	MA = Relación espúria
T es interviniente	I	PI = Contingencia	MI = Explicación o interpretación

(A. Almarcha, 1969, p. 133.)

Según este cuadro, se da especificación cuando la variable T es antecedente y aparece en uno de los parciales una relación entre X e Y más significativa que la inicial. Aunque T sea antecedente, si en los parciales no aparece relación entre X e Y, entonces estamos en el caso de la relación espúria o falsa entre X e Y. La contingencia existe si la variable es interviniente, posterior en el tiempo a X y anterior a Y, y en uno de los parciales aparece una relación más significativa entre X e Y que en la tabla inicial.

Por último, el tipo de interpretación se caracteriza porque la variable T sigue siendo interviniente, más en los parciales no aparece asociación entre X e Y, pero sí entre una de estas variables y T, en los marginales.

PROCEDIMIENTO

Supongamos que de la tabulación efectuada en una encuesta se han obtenido las siguientes tablas:

Absentismo laboral según el estado civil.

Absentismo laboral	*Estado civil de las mujeres*				
	Casadas		*Solteras*		
Alto	55 %	110	32 %	90	200
Bajo	45 %	90	68 %	190	280
TOTALES	100	200	100	280	480

Absentismo laboral según estado civil y edad.

Absentismo	*Mayores*					*Jóvenes*				
	Casadas		*Solteras*			*Casadas*		*Solteras*		
Alto	75%	90	75%	30	120	25%	20	25%	60	80
Bajo	25%	30	25%	10	40	75%	60	75%	180	240
TOTALES	100	120	100	40	100	100	80	100	240	320

Estas son las tablas parciales, que no son otra cosa que la misma tabla original pero descompuesta o dividida en dos partes o tablas distintas. Por ello, teniendo las tablas parciales y no la tabla original, se puede recomponer ésta.

Como he indicado anteriormente las tablas marginales se llaman así porque se forman con los marginales o totales de las tablas parciales. Estas tablas son dos: una que relaciona la variable independiente con la variable test o prueba y otra que relaciona esta misma variable con la variable Y, o variable dependiente. La primera tabla marginal se forma con los marginales horizontales de las dos tablas parciales y la segunda con los marginales verticales.

En el caso supuesto tendríamos para la primera tabla marginal:

ESTADO CIVIL SEGUN LA EDAD

Estado civil	*Mayores*		*Jóvenes*		
Casadas	120	75%	80	25%	
Solteras	40	25%	240	75%	
TOTALES	160	100	320	100	480

La segunda tabla marginal sería la siguiente:

ABSENTISMO SEGUN LA EDAD

Absentismo	*Mayores*		*Jóvenes*	
Alto	120	75%	80	25%
Bajo	40	25%	240	75%

Todas las tablas expuestas son fundamentales en el análisis multivariable. Este se funda en ellas. La manera de efectuarlo no consiste en otra cosa que en determinar, mediante la comparación de las distintas tablas entre sí, las variaciones sufridas en la asociación inicial de las variables de la tabla original con motivo de su puesta en relación con la nueva variable test o prueba.

Por tanto, para efectuar el análisis multivariable, lo primero que hay que realizar es determinar la asociación en la tabla original entre las variables X e Y, dependiente e independiente, para lo cual se suele emplear por su sencillez el coeficiente epsilon, como he indicado, aunque también se podrían emplear otros coeficientes, como por ejemplo el Q de Yule y el φ.

Establecido el coeficiente de asociación de la tabla original, entre X e Y, se determina igualmente la nueva asociación entre estas variables que se derive de las tablas parciales.

A continuación se comparan los coeficientes de asociación obtenidos para ver si ha habido variación o no en la asociación que suponía la tabla original.

Para completar el análisis multivariable, es preciso finalmente comparar igualmente la asociación de la tabla original entre x e y con la asociación que se derive de las tablas marginales entr e x y t, y t e y.

Los resultados de estas comparaciones sirven para determinar el tipo de análisis multivariable que corresponde al caso examinado y para aclarar la relación existente entre las variables independiente y dependiente probadas en este análisis.

Expuesto todo lo anterior, vamos a aplicarlo al caso concreto de las tablas que hemos formado como ejercicio:

En la tabla original aparece una relación previa entre x e y: epsilon $55 - 32 = 23\%$.

Pero en las tablas parciales esa asociación desaparece $75 - 75 = 0\%$; y $25 - 25 = 0\%$ en las dos.

Por lo que hace a las tablas marginales, vemos que aparece una asociación muy fuerte entre la variable t, la edad, y la variable x, estado civil: $75 - 25 = 50\%$; e igualmente entre la variable t, y la variable y, absentismo: epsilon: $75 - 25 = 50\%$.

Además la variable t, es claramente antecedente respecto a la variable independiente, estado civil; la edad es anterior al estado civil en el tiempo.

En conclusión tenemos por una parte que en las tablas parciales desaparece la asociación inicial entre x e y, y que en las tablas marginales

aparece una asociación muy significativa entre dichas variables y la variable t, por separado.

Además, la variable t es antecedente.

En consecuencia, se trata de un caso de espuriedad en el que la asociación inicial de la tabla original entre x e y es falsa o aparente, y está determinada por la fuerte asociación de dichas variables con la variable test, o prueba, que es la que influye en ambas variables x e y, sin que éstas variables se influyan entre sí. Ver ejercicios 235 a 237.

ORIGEN Y PRESUPUESTOS DEL "PATH ANALYSIS"

El path analysis es una técnica matemática que procede del campo de las ciencias naturales. Su fundador es el biómetra Sewell Wright. Alrededor del año 1960 se extendió en los Estados Unidos su utilización al campo de las ciencias sociales y de la sociología, donde en la actualidad es cada vez más empleado.

Los presupuestos en que se basa el análisis de senderos, es decir, las condiciones que debe reunir un modelo para que sea aplicable al mismo este tipo de análisis son los siguientes:

1. Modelos o sistemas de variables cerrados o completos. En este tipo de modelos, como señala Duncan (1974, p. 117) cada variable dependiente debe ser considerada explícitamente como completamente determinada por alguna combinación de variables en el sistema. En los casos en que no se mantiene la determinación completa por las variables medidas, debe introducirse una variable residual no correlaciocionada con otras variables determinantes del modelo.

2. El modelo debe ser recursivo. Un modelo es recursivo cuando las relaciones entre las variables que lo forman son o se supone que son en todo caso asimétricas. Este tipo de modelo implica que dos variables no pueden ser recíprocamente causa y efecto una de otra.

3. El modelo debe ser lineal. Esto quiere decir que las relaciones que unen las variables que lo forman se deben poder representar por ecuaciones lineales, y, en el caso de que estas ecuaciones sean de otro tipo, se deben transformar en lineales.

4. Relaciones de causa-efecto entre las variables del modelo. El análisis de senderos es una técnica para el análisis de las estructuras causales, y por tanto exige como condición que todas las variables estén unidas o se suponga que están unidas por relaciones de causalidad.

5. Nivel de medida de intervalo o de razón. El análisis de senderos requiere para su empleo que las variables del modelo sean de tipo cuantitativo, continuas, y que sus valores formen por tanto escalas de intervalo o de razón. También se pueden emplear variables cualitativas pero a condición de dicotomizarlas y cuantificarlas según el algebra de Boole.

6. Las variables residuales, también llamadas errores, que representan bien los errores de medición o las variables que pueden influir en el sistema pero no están incluidas en él, se supone que no están correlacionadas entre sí y que ejercen una influencia aleatoria sólo sobre una variable del modelo y no sobre varias.

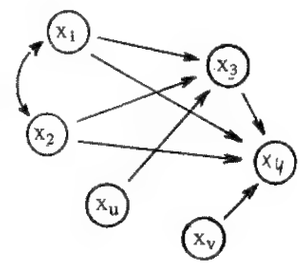
ELEMENTOS DEL ANALISIS DE SENDEROS

Los elementos básicos de este tipo de análisis son:

1. El diagrama de senderos.
2. El modelo de senderos.
3. Las ecuaciones estructurales.
4. Los coeficientes de Wright.

Veámoslos por separado.

1. El diagrama de senderos.— El diagrama de senderos (path diagram) es un dibujo mediante el cual se representan gráficamente las relaciones de causalidad que se supone existen en un conjunto de variables.



Como se ve en el diagrama adjunto, las variables se suelen representar por círculos que rodean la notación de la variable, y las relaciones entre las variables mediante flechas unidireccionales, rectas, que empiezan en la variable independiente que influye y cuya punta termina en la variable dependiente o influida.

La posible correlación entre las variables exógenas, por las que se entiende las no dependientes de otras variables del modelo, se suele representar por flechas de doble punta y con la línea de unión curva en lugar de recta. La posible relación entre estas variables no se tiene en cuenta ni entra a formar parte de las operaciones del análisis de senderos.

En el diagrama diseñado se pueden distinguir tres tipos de variables del modelo: endógenas, exógenas, y residuales. Las primeras son las variables x_3 y x_4 que dependen o figuran en el modelo influídas por otras variables del mismo. Exógenas son las variables x_1 y x_2 que son exter-

nas en cierto modo al modelo en cuanto no dependen de ninguna otra variable de las que lo forman. Residuales son las variables x_u y x_v que representan a los factores no observados, implícitos, que es posible influyan en cada una de las variables del modelo, así como a los posibles errores cometidos al medir las variables del modelo.

Las cifras que suelen figurar en los diagramas de senderos suelen ser, en el caso de las flechas orientadas hacia las variables endógenas, los coeficientes de Wright correspondientes a la relación entre las dos variables en cuestión, y, en el caso de las variables exógenas, los coeficientes de correlación simples de orden cero.

EL MODELO DE SENDEROS Y LAS ECUACIONES ESTRUCTURALES

Cualquier diagrama de senderos se puede transformar en un conjunto de ecuaciones que reciben el nombre de estructurales y que en su conjunto se suelen llamar modelo de senderos.

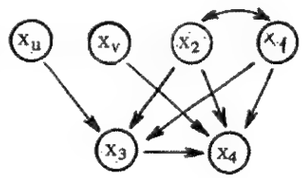
Como escribe Van de Geer (1971, p. 115), cualquier variable que según el diagrama sea dependiente de otra o de otras variables por acabar en ella una o más puntas de flecha, se puede expresar como una función, que, en este caso, ha de ser una ecuación lineal por exigencias del análisis, de las variables de las cuales parten las flechas.

Así, en el diagrama de senderos dado como ejemplo, x_3 y x_4 se pueden expresar como funciones de las variables x_1, x_2, x_u y x_1, x_2, x_3, x_v, respectivamente, dando lugar a las siguientes ecuaciones:

$$x_3 = p_{31}\, x_1 + p_{32}\, x_2 + p_{3u}\, x_u$$

$$x_4 = p_{41}\, x_1 + p_{42}\, x_2 + p_{43}\, x_3 + p_{4v}\, x_v$$

En las ecuaciones estructurales del análisis de senderos, las x_i representan los valores de las variables de la ecuación y los parámetros son los llamados coeficientes de senderos de Wright.



Estas ecuaciones, tratadas matemáticamente, se convierten en las ecuaciones estructurales.

COEFICIENTES DE WRIGHT

En las ecuaciones estructurales del análisis de senderos, los parámetros reciben el nombre de coeficiente de Wright. Como en el análisis de regresión, son estos coeficientes de Wright los que constituyen las incógnitas cuyo valor se halla mediante la resolución del sistema de ecuaciones estructurales del modelo.

El análisis de senderos consiste fundamentalmente, por tanto, en la determinación de estos coeficientes.

La notación de estos coeficientes suele ser p_{ij}. Los subíndices representan, el primero i la variable dependiente, y el segundo j la variable independiente.

Los coeficientes de Wright, p_{ij}, representan, como escribe Leik y Meeker (1975, p. 114) la relativa influencia de las diferentes variables sin contaminación de las varianzas producidas por otras variables.

Estos coeficientes coinciden con los coeficientes betha del análisis de regresión.

EFECTOS TOTAL, DIRECTO E INDIRECTO

También se pueden distinguir en el análisis de senderos los efectos total, directo y total indirecto.

El efecto total es la correlación de orden cero entre dos variables. Indica el efecto conjunto de una variable sobre otra variable dependiente de ella a través de todos los caminos directos e indirectos posibles.

El efecto directo es igual al coeficiente de Wright p_{ij} y el indirecto es la diferencia entre el coeficiente de correlación cero y el coeficiente de Wright.

En el caso de las variables x_1 y x_2 el efecto total indirecto, ETI, sería:

$$\text{ETI} = r_{12} - p_{12}$$

Si interesa saber detalladamente todas las influencias o efectos directos e indirectos de una variable sobre otra se desarrolla la ecuación del análisis de senderos que liga ambas variables en función únicamente de los coeficientes Wright, para lo cual se sustituyen los términos r_{ij} que entren en ella por los p_{ij} que correspondan a los mismos.

INTERPRETACION

Los resultados de este análisis se concretan en los coeficientes Wright obtenidos, que se refieren, como se ha visto, no sólo a las variables explícitas del modelo sino también a las implícitas (errores) y a la posibilidad de determinar los efectos directos e indirectos de cada variable independiente de una ecuación sobre las variables dependientes de ella.

Se basan, pues, en la naturaleza de los coeficientes obtenidos de regresión múltiple estandarizados y por tanto reflejan, en las variables explícitas, la influencia de la variable a la que se refieren en la variación de la variable dependiente, cuando las restantes variables del modelo se mantienen constantes.

En cuanto a los coeficientes Wright de los errores, su cuadrado representa la variación en la variable dependiente no explicada por las demás variables de la ecuación y que se atribuye a las variables implícitas o errores.

Por tanto, la diferencia entre 1 y el cuadrado de este coeficiente de Wright expresa la varianza en la variable dependiente de la ecuación explicada en su conjunto por las variables explícitas de la ecuación en cuestión.

Teniendo en cuenta esta significación de los coeficientes de Wright, la interpretación de los mismos puede consistir:

a) en el comentario de los efectos directos e indirectos de cada variable en las variables dependientes;

b) en el estudio de la influencia comparada de las distintas variables del sistema;

c) en la comparación de estos coeficientes, como coeficientes de correlación parcial que son, con los coeficentes de correlación total o de orden cero; y finalmente,

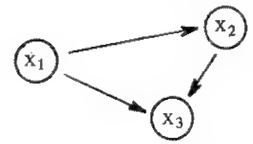
d) en la consideración de las varianzas explicadas en cada caso.

EJEMPLO NUMERICO DE ANALISIS DE SENDEROS

Sean las tres variables siguientes: renta, x_1, educación, x_2, y prestigio, x_3.

Supongamos que con base en la experiencia y en los conocimientos teóricos que tengamos postulemos que la renta influye en la educación y el prestigio, y que la educación influye a su vez en el prestigio social.

El diagrama que representaría este modelo sería, pues, el siguiente:



Aplicando la regla expuesta de que cualquier variable, que según el diagrama sea dependiente de otra u otras por acabar en ella una o más puntas de flecha, se puede expresar como una función de las variables de las cuales parten las flechas, se tiene en este caso primero que en la variable x_1 no acaba ninguna punta de flecha, por tanto no se puede formar ecuación respecto de la misma. En cambio en el caso de la variable x_2 acaba una punta de flecha, la de la variable x_1. Por tanto la variable x_2 es dependiente de la x_1, y se puede formar la siguiente ecuación:

$$x_2 = p_{12}\ x_1$$

Por el mismo motivo, la variable x_3, al acabar en ella dos puntas de flecha, las de las variables x_1 y x_2, depende de estas dos ecuaciones variables, lo que se expresa mediante esta segunda ecuación:

$$x_3 = p_{31}\ x_1 + p_{32}\ x_2$$

Multiplicando estas dos ecuaciones por todas las variables inferiores al índice menos 1 de la variable a que se refieren, en concreto, en este ejemplo, la ecuación sobre x_2 por x_1 y la referente a x_3 por x_2 y x_1, sustituyendo en estas ecuaciones multiplicadas las covarianzas $(x_i\ x_j)$ por los coeficientes r de Pearson, se obtienen las ecuaciones estructurales del análisis de senderos siguientes:

$$r_{31} = p_{31} + p_{32}\ r_{12}$$
$$r_{32} = p_{32} + p_{31}\ r_{12}$$
$$r_{12} = p_{12}$$

Supóngase ahora que la matriz de correlaciones entre las variables en cuestión es la siguiente:

	x_1	x_2	x_3
x_1	—	.70	.60
x_2		—	*.75*
x_3			—

Sustituyendo en las ecuaciones estructurales, los coeficientes r_{ij} de Pearson por su valor según la matriz anterior, se llega a dos sistemas de ecuación: uno de una sola ecuación y una sola incógnita que tiene solución directa, y otro de dos ecuaciones con dos incógnitas.

Primer sistema:

$$p_{21} = .70$$

Segundo sistema:

$$.60 = p_{31} + .70\, p_{32}$$
$$.75 = p_{32} + .70\, p_{31}$$

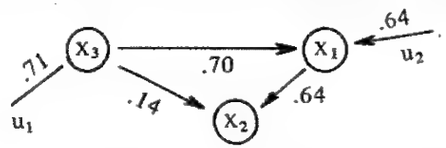
Resolviendo este último sistema se tienen los siguientes valores de los coeficientes p_{ij} de Wright buscados:

$$p_{21} = .70; \qquad p_{32} = .64; \qquad p_{31} = .14$$

Por último se aplicaría la fórmula correspondiente para hallar los p_{iu} de los términos residuales, con este resultado:

$$p_{3u} = \sqrt{1 - p_{31}\, r_{31} - p_{32}\, r_{32}} = \sqrt{1 - .098 - .48} = .645;$$
$$p_{2u} = \sqrt{1 - p_{21}\, r_{21}} = \sqrt{1 - .70\,(.70)} = .71$$

Con los resultados se puede formar el siguiente diagrama de senderos:



De todo ello se deriva que en este modelo existe una apreciable influencia de la educación sobre la renta y de ésta sobre el prestigio y una influencia reducida de la educación sobre el prestigio.

COEFICIENTES d

Constituyen, en conjunto, un procedimiento para el análisis de las tablas de contingencia, ideado por J.A. Davis, basado en las diferencias de proporciones entre las casillas de la tabla, en el que se combina la lógica del análisis multivariable de Lazarsfeld y del procedimiento de elaboración de Rosenberg con la representación de las relaciones entre las variables mediante

los llamados en inglés "flow—graph" y su expresión matemática por sistemas de ecuaciones lineales.

Específicamente reciben el nombre de coeficientes d, las diferencias entre las proporciones de las distintas casillas de una tabla de contingencia que, separadamente, indican la probabilidad diferencial de estar en una casilla y no en otra que se toma como base de comparación. Estos coeficientes se caracterizan por ser asimétricos, poder variar de +1 a —1 y tener una significación similar a los coeficientes *a* y *b* del análisis de regresión. La lógica indicada del análisis basado en estos c. consiste en probar la relación inicial entre dos variables mediante la introducción sucesiva y el control de nuevas variables. Este control se realiza mediante los procedimientos de control de una tercera variable o de estandarización. De modo similar al análisis de senderos, en este análisis, las relaciones que existen entre las variables se representan mediante grafos, formados por flechas, que van de las variables que se cree influyen a las influidas; y para formular las ecuaciones correspondientes se tiene en cuenta que, linealmente, cada variable dependiente, determinada por una o más variables independientes es igual a la suma de los productos de las variables independientes por los coeficientes d correspondientes, más una constante k o coeficiente *b* del análisis de regresión, que es igual al valor de la variable dependiente cuando la variable independiente es cero. Además, en este análisis, se estudian también los efectos indirectos, o de una variable independiente en la dependiente a través de otra o de otras independientes, y la existencia o no de interacción entre las variables independientes en su influjo sobre la variable dependiente. (V. ejercicio 237 a).

Objetivos de este análisis, y en general a los que se realizan con tres variables, son (Sánchez Carrión 1975, 286):

1. Descubrir si la relación entre dos variables es causal o meramente estadística.

2. Especificar en la relación causal entre dos variables cuál de ellas es la variable independiente y cuál la dependiente.

3. Poner de manifiesto la relación verdadera entre dos variables, velada por la intervención en ella de una tercera variable.

4. Determinar las modalidades de la relación entre dos variables cuando se tienen en cuenta las distintas categorías de ambas.

5. Ver el efecto conjunto no ya de una variable sino de dos en una tercera.

EJERCICIOS

235

En una investigación sobre mercados se buscó la relación entre la publicidad y la adquisición de un aparato adelgazante. Después de establecer la tabla de los resultados obtenidos se hizo intervenir otra variable, el sexo, y se obtuvo la siguiente tabla:

	SEXO					
	VARONES			MUJERES		
	PUBLICIDAD			PUBLICIDAD		
	Sí	*No*		*Sí*	*No*	
Adquirieron el aparato	40 (40)	20 (10)	50	70 (150)	25 (60)	210
No lo adquirieron	60 (60)	80 (40)	100	30 (45)	75 (180)	225
	100 (100)	(50)	150	(195)	(240)	435

a) Establecer la tabla original y los épsilons de todas ellas.

b) Determinar si la nueva variable introducida es antecedente o interviniente y establecer el gráfico del modelo correspondiente.

c) Determinar el tipo de relaciones entre variables obtenido al introducir la nueva variable, sexo.

RESPUESTA

a) Los épsilons de la tabla dada son 20 por 100 y 45 por 100.

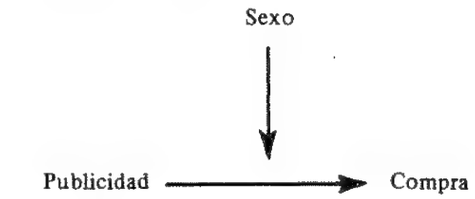
Para formar la tabla original hay que sumar los valores absolutos correspondientes que figuran en las dos tablas de: adquirieron el aparato y publicidad, sí, y adquirieron el aparato, no e igual en los que no lo adquirieron, y después hallar los porcentajes de cada casilla:

	Publicidad			
	Sí (%)		*No* (%)	
Adquisición del aparato	190	64	70	24
No adquisición del aparato	105	36	220	76
	295	100	290	100

Epsilon de esta tabla = 40 por 100.

2. La variable introducida, sexo, es antecedente respecto a la publicidad, en cuanto es anterior en el tiempo a ésta.

El esquema del modelo será, pues:



3. La comparación del épsilon de la tabla original, 42, con los épsilons de las tablas formadas con la introducción de la nueva variable, sexo, 20 y 45, muestra que, mientras en una de las tablas parciales la de las mujeres, la relación entre las variables publicidad y adquisición del aparato permanece sustancialmente incambiada, 40 y 45 por 100, de épsilons, en la otra tabla parcial, la de los varones, aparece una asociación significativamente más reducida, épsilons 20 y 40. Por tanto, al ser la variable t antecedente, estamos ante un caso de especificación en el que la introducción de la variable test o prueba sexo especifica la relación original en el sentido de mostrar que es en el sexo femenino donde la relación entre publicidad y compra es más estrecha. Como lo indica el modelo de relación entre las variables la compra o no compra del aparato depende de la publicidad, y esta relación depende del sexo.

236

En una investigación de sociología política se halló primero la relación entre actitud política y el estado civil. Después de establecer la relación se introdujo la variable edad y se obtuvo la siguiente tabla.

	EDAD				
	MAYORES DE 35 AÑOS				
	ESTADO CIVIL				
Filiación política	*Casado*		*Soltero*		
Monárquico	75 %	(120)	75 %	(30)	150
Republicano	25 %	(40)	25 %	(10)	50
		160		40	200

	EDAD			
	DE 35 AÑOS O MENOS			
	ESTADO CIVIL			
Filiación política	*Casado*		*Soltero*	
Monárquico	40 %	(40)	40 % (80)	120
Repúblicano	60 %	(60)	60 % (120)	180
		100	200	300

1.

a) Establecer la tabla original.

b) Formar con base en los marginales las tablas de relación de la edad con la filiación política y el estado civil.

c) Interpretar los resultados realizando el análisis multivariable, con determinación del tipo de resultados obtenidos, especificación, contingencia, espureidad o explicación.

d) Establecer si la variable introducida es antecedente o interviniente y diseñar el gráfico del modelo correspondiente.

RESPUESTA

Tabla original:

	Estado civil			
	Casado		*Soltero*	
Filiación política:				
Monárquico	62 %	160	46 %	110
Republicano	38 %	100	54 %	130
TOTALES	100 %	260	100 %	240

b) Tablas marginales:

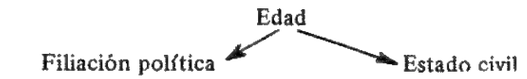
	Edad			
	Más de 35 años		*35 o menos*	
Filiación política:				
Monárquico	75 %	(150)	40 %	(120)
Republicano	25 %	(50)	60 %	(180)
TOTALES	100 %	(200)	100 %	(300)

	Edad			
	Más de 35 años		*35 o menos*	
Estado civil:				
Casado	80 %	(160)	33,3 %	(100)
Soltero	20 %	(40)	66,6 %	(200)
TOTALES	100 %	(200)	100 %	(300)

c) Interpretación:

Según la tabla original, existe una cierta asociación positiva, ϵ 14, entre el estado civil y la actitud política. En las tablas parciales esta asociación desaparece. Por el contrario, las tablas marginales muestran una relación mayor a la de la tabla original entre la edad y la filiación política, épsilon, 35 por 100, y entre la edad y el estado civil, épsilon, 46,6 por 100. Se trata, pues, de una relación espúria en la tabla original debida a la fuerte relación de las dos variables que la forman con la variable test o prueba.

d) La variable introducida es antecedente, ya que la edad es anterior al estado civil. Ello confirma la calificación de espureidad, pues se dan las dos condiciones de esta categoría, al carácter antecedente de la variable test y la ausencia de asociación en las tablas parciales. El diseño del modelo sería, pues, el siguiente:



237

1. Dada la siguiente tabla:

	PADRE EN PARO				
Hijas prostitutas	*Sí*		*No*		
Sí	71%	285	36%	218	400
No.....	29%	115	64%	382	600
TOTALES	100%	400	100%	600	1.000

Hallar épsilon e interpretar los resultados.

RESPUESTA

Epsilon = 35 por 100. Existe una relación positiva según la tabla entre la situación de paro del padre y la prostitución de las hijas.

2. Las variables de la tabla anterior se han relacionado con la desorganización familiar (alta o baja) y han resultado las siguientes tablas:

Se pide:

a) Calcular épsilon, indicar el carácter de la variable test o prueba introducida e interpretar los resultados obtenidos en relación a la tabla original.

b) Señalar si es necesario formar las tablas de los marginales para comprobar la transformación producida por la introducción de la variable test y en caso afirmativo formar dichas tablas.

c) Realizar la explicación general de los resultados en términos del análisis multivariable y formar el modelo de la relación entre las variables.

DESORGANIZACION FAMILIAR

ALTA

Prostitución de las hijas	*Padre en paro*		*Padre ocupado*		
Sí	90%	(270)	85%	(170)	440
No	10%	(30)	15%	(30)	60
SUMAS	100%	(300)	100%	(200)	500

DESORGANIZACION FAMILIAR

BAJA

Prostitución de las hijas	*Padre en paro*		*Padre ocupado*		
Sí	15%	(15)	12%	(48)	63
No	85%	(85)	88%	(352)	437
SUMAS	100%	(100)	100%	(400)	500

RESPUESTA

a) Epsilon 5 y 3 por 100. La variable introducida se puede considerar interviniente o posterior a la otra variable independiente, el paro. La interpretación de los resultados muestra que la relación en la tabla original, épsilon 35 por 100, desaparece prácticamente en las parciales, épsilon 5 y 3 por 100. Por tanto, el tratarse de una variable test interviniente se está en el caso de la explicación o interpretación.

b) Al no aparecer una relación significativa en las tablas parciales se debe formar las tablas de los marginales para comprobar si en éstas aparece asociación.

c) Se ve que mientras el épsilon de la desorganización familiar y el paro, 40 por 100, asciende en relación a la tabla original, 35 por 100; el de la tabla desorganización familiar y prostitución aumenta significativamente al pasar del 35 al 66 por 100. Por tanto, se verifica la condición de la explicación de aparecer una relación más

Hijas prostitutas	DESORGANIZACION FAMILIAR *Alta*		DESORGANIZACION FAMILIAR *Baja*		*Paro*	DESORGANIZACION FAMILIAR *Alta*		DESORGANIZACION FAMILIAR *Baja*	
Sí	440	88 %	63	12 %	Sí	300	60 %	100	20 %
No	60	12 %	437	88 %	No	200	40 %	400	80 %
TOTAL . . .	500	100 %	500	100 %	TOTAL . .	500	100 %	500	100 %

significativa en una de las tablas marginales. En conclusión, es la variable test, la desorganización familiar la que explica la relación existente en la tabla original entre el paro y la prostitución de las hijas, como lo confirma el hecho de que al controlar o mantener constante en las tablas parciales la desorganización familiar, la relación inicial entre el paro y la prostitución de la hijas casi desaparece.

La proporción de hijas prostitutas es mayor en el caso de padres en paro, porque la desorganización familiar es más elevada.

El modelo será: Paro → Desorganización familiar → Prostitución de las hijas.

237 a.

En la tabla de contingencia siguiente, hallar los coeficientes d

Renta	Estudios	Absentismo Bajo	Absentismo Alto
Baja	No superiores	400	300
Baja	Superiores	100	10
Alta	No superiores	250	150
Alta	Superiores	50	5

RESPUESTA

Para formar el grafo se puede partir del supuesto de que la renta influye en el nivel de estudios y que ambas influyen en el absentismo:

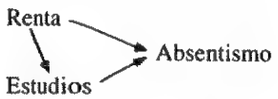
Se puede comprobar primero la asociación o no entre la renta y los estudios:

Renta	Estudios		
	No supers.	Supers.	Total
Baja	700	110	810
Alta	400	55	455
Total	1.100	165	1.265

Tabla de proporciones: (Se halla en la dirección de la variable independiente, la renta, y las diferencias entre ellas, o coeficientes d, en la dirección contraria).

Renta	Estudios		
	—	+	Total
—	.86	.14	1.
+	.87	.13	1.
c.d	—.01	+.01	

Los c.d. obtenidos son mínimos y por tanto indican la no existencia de relación entre ambas variables, con lo que el grafo inicial resulta ser, según los datos, el siguiente:



Para determinar el efecto de la renta y el nivel de estudios, separadamente, en el absolutismo, se forman las siguientes tablas, en cifras absolutas y proporciones:

Estudios	Absentismo				
	—	+	T.	—	+
No superiores	650	450	1.100	.59	.41
Superiores	150	15	165	.91	.09
Coeficientes d				—.32	+.32

Renta	Absentismo				
	—	+	T.	—	+
Baja	500	310	810	.62	.38
Alta	300	155	455	.65	.35
Coeficientes d				—.03	+.03

Estas tablas muestran un efecto positivo del nivel de estudio en el absentismo y en efecto también positivo, pero insignificante de la renta en el absentismo.

Para comprobar la no existencia de interacción en este caso, en cuanto, según los datos, no existe asociación entre las variables renta y estudios, se puede emplear la estandarización o como se hace seguidamente, obtener los d condicionales, que se refieren a la relación entre la renta y el nivel de estudios condicionada o relativa a una sola categoría del absentismo.

Estudios	Absentismo bajo				
	Renta				
	Baja	Alta	Baja	Alta	d condicionales
No superiores	400	250	.57	.62	+.05
Superiores	100	50	.91	.91	0
d condicionales			+.34	+.29	

La diferencia entre los d condicionales tan reducida, 0.05, confirma la no existencia de interacción entre las variables, renta y nivel de estudios. La proporción .57, es el resultado de dividir los 400, con estudios superiores y renta baja, de absentismo bajo, con el total de absentismo bajo y alto, 700, y así sucesivamente.

31. ANALISIS DEL CAMBIO SOCIAL

Así como el análisis causal se basa en la noción de causa, el análisis del cambio social se basa en los conceptos de cambio y de proceso social.

Cambio social, como la misma palabra cambio, es un término general que abarca toda variación, modificación, transformación, etc. en los sujetos sociales o en sus elementos o características.

En el cambio social, como en todo cambio, se pueden distinguir un estado inicial, un estado final, y un tránsito del primero al segundo.

Por ello todo cambio social implica un proceso, entendida esta palabra como la sucesión de las fases o etapas que supone el paso de la situación inicial a la final.

El proceso social, según esto, no es otra cosa que un aspecto o elemento del cambio social y representa la forma en que tiene lugar el cambio o paso de un estado a otro.

El cambio social se puede estudiar, de acuerdo con los aspectos distinguidos en él:

a) En el aspecto inicial del cambio con referencia a las causas y factores que lo producen e intervienen en el cambio;

b) En su proceso de paso de la situación inicial a la final; y

c) En la situación alcanzada a consecuencia del cambio.

El estudio de la situación inicial y de las causas del cambio, así como de la situación final, corresponden más bien al estudio de los que se han llamado, haciendo uso de la abstracción pues en la realidad no existen propiamente situaciones estáticas, modelos estáticos, así como a los

modelos causales, que son los que estudian los efectos producidos por las causas.

Queda, pues, el estudio del proceso que supone todo cambio social. De ahí que, desde este punto de vista se pueda decir que el estudio del cambio social en general consiste, en último término, en el análisis de los procesos social.

En este sentido, Hamblin, Jacobsen y Miller (1973, p. 1) consideran el cambio social en términos de los procesos cuantitativos que ocurren a lo largo del tiempo, y construyen su teoría alrededor de proposiciones generales o ecuaciones que describen la forma de estos procesos.

Los modelos para el análisis del cambio social pueden ser muy diversos, pero aquí únicamente se puede hacer referencia, al análisis simple de series temporales y al análisis de las tablas de rotación.

EL ANALISIS ESTADISTICO DE LAS SERIES TEMPORALES

Una serie temporal está formada por un conjunto de valores obtenidos de observaciones referentes al mismo fenómeno, realizadas en una sucesión de momentos de tiempo, normalmente a intervalos iguales.

Este tipo de análisis, aunque no sea tan completo como si se emplearan modelos matemáticos más complejos, sin embargo, es frecuentemente más práctico, dada la dificultad que presentan los modelos matemáticos en las ciencias sociales y su frecuente escasa exactitud y dificultad de interpretación.

En este tipo de análisis se estudian los resultados de la evolución de los fenómenos sociales, haciendo abstracción de las causas que hayan producido esta evolución y de las variables que hayan intervenido en la misma.

El punto de partida de este análisis en su representación gráfica, que se realiza simplemente llevando en un eje de coordenadas los intervalos de tiempo al eje de las X, y los valores del fenómeno en cada momento del tiempo al eje de las Y.

La observación repetida de muchas representaciones gráficas de series cronológicas ha permitido distinguir en el cambio o evolución de los fenómenos, cuatro elementos, aspectos o tipos de variación principales:

a) Un componente de larga duración, secular, también llamado de tendencia.

b) Otro componente cíclico, de duración media, constituido por oscilaciones alrededor de la tendencia central.

c) Otro componente estacional formado por variaciones cortas que se repiten periódicamente con regularidad similar a las de las estaciones del año solar.

d) Finalmente, un componente irregular, formado por variaciones ocasionales y aleatorias que parecen no obedecer a norma alguna detectable.

Estos componentes se corresponden con los tipos de factores que se pueden distinguir lógicamente en la evolución y cambio de los fenómenos.

Entre estos factores, normalmente hay algunos persistentes, cuyo influjo permanece durante largos períodos de tiempo; otros son de alcance medio, los cíclicos; otros actúan según intervalos de tiempo más cortos y regulares; y, finalmente, están los factores imprevistos y aleatorios sin regularidad observable alguna.

El análisis estadístico de las series temporales se concreta, precisamente, en las distintas técnicas que existen para el estudio estadístico de los distintos elementos citados, que se puede distinguir en las series cronológicas. De ellas, se expone únicamente, como procedimiento básico de análisis de una serie temporal, el empleo del análisis de regresión para la estimación de la tendencia de una serie temporal.

FUNDAMENTOS TEORICOS DEL MODELO

No se trata de un modelo causal. Realmente el tiempo no es causa de nada, porque no produce por sí, directamente, nada, aunque es presupuesto de todo, en cuanto interviene o está siempre presente como una condición en todo proceso causal.

Los fenómenos evolucionan a lo largo del tiempo, pero dicha evolución no es causada propiamente por el tiempo, sino por diversos factores a veces desconocidos. No obstante, haciendo caso omiso de los factores que producen la evolución del fenómeno, se puede estudiar matemáticamente esta evolución en función del tiempo.

Ontológicamente, constituyen, pues, modelos funcionales y no causales e indican la correspondencia entre la variación del fenómeno investigado y la variación del tiempo.

El fundamento teórico de este modelo se encuentra, pues, en la evolución de los fenómenos, en este caso sociales, a lo largo del tiempo.

Aunque esta evolución no sea causada propiamente por el tiempo sino por otros factores como se ha indicado, el tiempo está siempre presente en ella y es claro que existe una correspondencia entre el paso del tiempo y la evolución o cambio de los fenómenos.

De ahí que sea posible siempre en todo fenómeno, estudiar la relación que guarda la evolución del fenómeno en cuestión con el proceso de paso del tiempo, considerado éste como condición o variable independiente de la transformación del fenómeno de que se trate.

REPRESENTACION GRAFICA DEL MODELO Y ESPECIFICACION DE SUS ECUACIONES MATEMATICAS

La representación gráfica no presenta ninguna dificultad. Está constituida por una simple flecha que pone en relación la variable tiempo con la variable referente al fenómeno estudiado, orientada en el sentido de esta última como variable dependiente.

$$T \longrightarrow Y$$

Para la especificación de la única ecuación del modelo, ya que en este caso no está especificado por la fundamentación teórica, y que puede ser muy diversa según los fenómenos estudiados, será conveniente diseñar la huella empírica del fenómeno y decidir por la observación de ésta, la forma de la ecuación que creemos más se ajusta a los datos.

Entre las diferentes ecuaciones posibles son corrientes la lineal:

$$Y = a + bT;$$

la parabólica:

$$Y_t = a + bT + cT^2;$$

la exponencial:

$$Y = a + b \log T.$$

OBTENCION DE LOS DATOS EMPIRICOS SOBRE LAS VARIABLES DEL MODELO

La obtención de los datos empíricos sobre las variables de los modelos de este tipo, no ofrece, por lo general, dificultad, sobre todo cuando se trata del estudio de la evolución temporal de variables demográficas, macrosociológicas (emigración, movilidad social, etc.), macroeconómicas, etc.

Normalmente se encuentran a disposición de los investigadores, mediante las estadísticas sobre dichas variables que se realizan y publican periódicamente en todos los países. Lo único que se requiere es la utilización de dichas estadísticas con sentido crítico y procurando asegurarse antes del grado de fiabilidad que merecen.

Para la especificación de la ecuación del modelo y hallazgo de los parámetros de dicha ecuación, el procedimiento es el del análisis de regresión (Ver ejercicio 239).

TABLAS DE ROTACION

Son tablas de contingencia cuadradas, con el mismo número de filas y columnas, referentes ambas a las mismas categorías de una o más variables determinadas. Estas categorías iguales de las filas y columnas pueden referirse a los mismos sujetos en dos tiempos distintos o a distintos sujetos, que representan también tiempos diversos, como es el caso de la movilidad intergeneracional o movilidad de los hijos en su ocupación respecto a la que tenían sus padres. Las frecuencias de sus casillas tienen el mismo significado, en general, que el indicado de las tablas de contingencia.

En las tablas de rotación se pueden distinguir dos tipos básicos:

a) Las tablas de rotación en las que se relacionan las categorías de una sola variable, con dos o más categorías, en dos tiempos distintos, y

b) Las tablas de rotación en las que se relacionan dos variables dicotómicas y no una sola como antes en dos momentos sucesivos del tiempo.

Como las tablas de rotación relacionan la variable o variables a que se refieren en dos momentos diferentes del tiempo, su análisis permite

estudiar los procesos sociales a los que se apliquen o la evolución de un fenómeno social de un momento a otro del tiempo.

Este análisis, por tanto, tiene muchas aplicaciones, como, por ejemplo, en demografía para investigar las variaciones en la estructura o composición de la población de un tiempo a otro; en sociología electoral para el estudio de los cambios en el voto a los partidos políticos de una elección a otra; y, en fin, en la investigación de panel para el examen de las transformaciones producidas en un grupo con el transcurso del tiempo, cuando este grupo es objeto de dos encuestas distintas separadas por un intervalo de tiempo determinado.

A continuación, se trata, como ejemplo del análisis de las tablas de rotación de una sola variable y de las de dos con dos categorías: del análisis de las tablas de movilidad social y del de las tablas de dieciséis casillas.

TABLAS DE ROTACION DE LA MOVILIDAD SOCIAL

Una de las divisiones de la movilidad social es la que distingue entre movilidad social intrageneracional y movilidad social intergeneracional.

La primera se refire al cambio de ocupación de los mismos sujetos de un tiempo dado t_1 a otro t_2. Por el contrario, la segunda atañe al cambio de ocupación de los hijos respecto a la ocupación de sus padres. padres.

Las tablas de movilidad social se pueden referir a ambas clases de movilidad social indistintamente. La diferencia consiste en que, en el caso de la movilidad intrageneracional, las cabeceras de las filas de la tabla representan las ocupaciones de los individuos en el tiempo uno, mientras que en el de la movilidad intergeneracional, representan la ocupación del padre.

Similarmente, las cabeceras de las columnas de la tabla representan la ocupación de los sujetos en el momento t_2, cuando se trata de movilidad intrageneracional, y la ocupación de los hijos cuando se trata de movilidad intergeneracional.

Supuesto esto, la tabla de movilidad social es una tabla de contingencias clasica, o de doble entrada, en la que en cada casilla de la tabla se consigna el número de casos que reúne a la vez las dos dimensiones o categorías que se cruzan en cada casilla.

Así, la casilla de la fila 2 y la columna 4 nos dice, en el caso de la movilidad intrageneracional, el número de sujetos que teniendo la ocupación 2 en el tiempo t_1 han pasado a la ocupación 4 en el tiempo t_2.

En el caso de la movilidad intergeneracional, la misma casilla contiene el número de parejas de padres e hijos, en las que el padre tiene la ocupación 2 y el hijo la ocupación 4.

La notación genérica de estas tablas sigue las reglas de la notación matricial. Las filas y las columnas se numeran correlativamente a partir del 1, de arriba a abajo las filas y de izquierda a derecha las columnas, y a cada casilla se le designa con la letra n, y dos subíndices, el primero el número de la fila y el segundo el de la columna. Los totales o marginales de las filas se designan igualmente con la letra n y con el número de la fila como primer subíndice y un punto como segundo subíndice.

Por su parte, los totales de las columnas se designan también con la letra n, pero con un punto como primer subíndice.

El total general de la tabla se representa por la N mayúscula.

Las tablas de movilidad social constituyen una matriz cuadrada, con un número de columnas y filas igual al número de ocupaciones que se distinguen y su notación en conjunto adopta la forma genérica siguiente:

		T_2 Ocupaciones				
	Ocupaciones	*1*	*2*	...	*x*	Σ
	1	n_{11}	n_{12}	...	n_{1x}	$n_{1.}$
	2	n_{21}	n_{22}	...	n_{2x}	$n_{2.}$
T_1	⋮	⋮	⋮		⋮	⋮
	x	n_{x1}	n_{x2}	...	n_{xx}	$n_{x.}$
	Σ	$n_{.1}$	$n_{.2}$	...	$n_{.x}$	N

Respecto a esta tabla básica podemos hallar la proporción que representa cada frecuencia de las casillas de la tabla, respecto a N, total general de la tabla, y obtener así la matriz de movilidad social en proporciones siguientes, en la que su suma o total general será igual a uno. En ella $p_{ij} = n_{ij}/N$.

		T_2 Ocupaciones			
	Ocupaciones	*1*	*2*	*x*	Σ
T_1	*1*	p_{11}	p_{12} ...	p_{1x}	$p_{1.}$
	2	p_{21}	p_{22} ...	p_{2x}	$p_{2.}$
	⋮	⋮	⋮	⋮	⋮
	x	p_{x1}	p_{x2}	p_{xx}	$p_{x.}$
	Σ	$p_{.1}$	$p_{.2}$	$p_{.x}$	1

Igualmente se puede dividir o bien los p_{ij} por el total $p_{i.}$ de cada fila, o bien los n_{ij} por el total $n_{i.}$ de cada fila, lo que es lo mismo, obteniendo así las proporciones condicionales, y formar con ellas otra nueva matriz de movilidad social que recibe el nombre de transición.

En esta casilla la suma de cada fila p será igual a la unidad.

	Ocupaciones				
Ocupaciones	*1*	*2*	...	*x*	Σ
1	$p_{1,1}$	$p_{1,2}$	...	$p_{1,x}$	1
2	$p_{2,1}$	$p_{2,2}$	...	$p_{2,x}$	1
⋮	⋮	⋮		⋮	⋮
x	$p_{x,1}$	$p_{x,2}$	...	$p_{x,x}$	1

ANALISIS DESCRIPTIVO DE LAS TABLAS DE MOVILIDAD SOCIAL

Las tablas de movilidad social descritas reciben el nombre de movilidad social bruta. Esta tabla se puede descomponer en otras tres que reciben el nombre de tablas de inmovilidad, tabla de cambio y tabla residual (Vid. Doreian, 1973, p. 114).

Sea la siguiente tabla de movilidad bruta:

		T_2			
		I	*II*	*III*	*T.*
T_1	*I*	100	50	80	230
	II	200	300	90	590
	III	250	190	400	840
	T.	550	540	570	1640

En ella podemos obtener la tabla de inmovilidad que está formada únicamente por las mismas frecuencias que figuran en la tabla de movilidad bruta para las casillas de la diagonal principal y por ceros en las restantes casillas. Se llama tabla de inmovilidad porque las frecuencias de las casillas de la diagonal principal representan el número de los que no han cambiado de ocupación del tiempo t_1 al t_2, o el número de los hijos que tienen la misma ocupación de los padres.

En el ejemplo expuesto esta tabla sería la siguiente:

		T_2		
		I	*II*	*III*
T_1	*I*	100		
	II		300	
	III			400

En segundo lugar, se puede formar la tabla de cambio en la que las casillas de la diagonal principal están vacías y en las restantes casillas, que son las que reflejan los cambios y se disponen simétricamente alrededor del eje de la diagonal principal, se consigna, en cada par de casillas simétricas, la frecuencia en que ambas casillas coinciden, que, en todo caso, es la menor de las que figuran en la tabla de movilidad bruta, respecto a dichas casillas.

En esta tabla, suponiendo que las ocupaciones están ordenadas de menor a mayor categoría, las frecuencias de las casillas a la derecha de la diagonal principal representan los cambios en dirección ascendente, los que han cambiado ascendiendo en la categoría de su ocupación ini-

cial, y las frecuencias de las casillas situadas a la izquierda de la diagonal principal representan los cambios en dirección descendente, es decir, los que han pasado a ocupar una ocupación más baja de la inicial o de la que tenían sus padres.

		T_2		
		I	*II*	*III*
T_1	*I*		50	80
	II	50		90
	III	80	90	

La tabla de cambio, implica que, cuando no son iguales las casillas simétricas en la matriz de movilidad bruta, hay un excedente en la mayor de ellas, que indica el exceso de cambio en el sentido que corresponde a la mayor frecuencia.

Si la diferencia en las casillas simétricas es cero indica que los cambios ascendentes y descendentes son iguales, pero si son distintas señalan que los cambios ascendentes superan a los descendentes en cada par de casillas simétricas o al revés.

La tabla de movilidad residual no es, pues, otra cosa que la formada por estas diferencias o excedentes en sus casillas correspondientes y por ceros en todas las demás.

En el ejemplo propuesto sería esta:

Tabla de movilidad residual

	T_1		
	I	*II*	*III*
I	0	0	0
II	150	0	0
III	170	110	0

Tabla de movilidad neta

	T_1	T_2	*Movil. neta*
I	230	550	+320
II	590	540	−50
III	840	570	−270

Finalmente de la tabla de movilidad bruta se puede derivar también la movilidad neta, con base en los marginales o totales de la tabla de movilidad bruta.

La movilidad neta respecto a cada ocupación está constituida por la diferencia entre los totales de la fila y la columna referentes a dicha ocupación.

En una matriz, la suma de las movilidades netas de todas las ocupaciones que forman la matriz debe ser igual a cero.

Cualquier matriz o tabla de movilidad social bruta se puede analizar descriptivamente hallando y estudiando comparativamente la importancia de sus elementos componentes indicados; inmovilidad, movilidad de cambio, movilidad residual y movilidad neta.

Este análisis se puede concretar en una serie de parámetros entre los cuales se pueden destacar:

a) El índice de movilidad social en la tabla, que en el ejemplo de la tabla de movilidad social 2 x 2 siguiente:

n_{11}	n_{12}
n_{21}	n_{22}

sería:

$$n_{12} + n_{21}$$

b) El índice de inmovilidad, igual en este caso a

$$n_{11} + n_{22}$$

c) El índice de movilidad ascendente, igual a n_{12}.

d) El índice de movilidad descendente, igual a n_{21}.

e) Indice de movilidad residual: total, ascendente y descendente que sería igual a la suma de todas las frecuencias halladas en la tabla el primero; a la suma de dichas frecuencias que se hallan a la derecha de la diagonal principal el segundo; y a la suma de dichas frecuencias a la izquierda de la diagonal principal el tercero.

Se entiende con referencia a la tabla residual.

(Ver ejercicios 240 a 242).

TABLA DE MOVILIDAD PERFECTA Y SUS INDICES

Otro instrumento que se puede utilizar para el estudio descriptivo de la movilidad social es la tabla de movilidad perfecta.

Desde el punto de vista sociológico, esta hipótesis de la movilidad perfecta implica que no existe ninguna restricción social que impida la movilidad de los individuos que forman el grupo estudiado.

En este supuesto, todos los miembros de grupo tienen la misma oportunidad de cambios de ocupación, y la posición de un sujeto del grupo en un momento dado t_i, no depende para nada de su posición anterior t_{i-1}.

Estadísticamente, este supuesto se produce en la tabla de contingencia que implica la tabla de movilidad social, cuando no hay asociación alguna entre las variables constituidas por la distribución de las ocupaciones en los tiempos t_1 y t_2.

Sea la siguiente tabla de movilidad bruta:

	I	*II*	*III*	
I	110	70	20	200
II	30	310	60	400
III	10	220	170	400
	150	600	250	1000

De acuerdo con lo anterior, la tabla de movilidad perfecta sería la siguiente:

	I	*II*	*III*	
I	30	120	50	200
II	60	240	100	400
III	60	240	100	400
	150	600	250	

Dadas ambas tablas, se puede para cada casilla de las mismas hallar la razón de movilidad, dividiendo simplemente la frecuencia empírica de cada casilla de la tabla obtenida en realidad por las frecuencias teóricas de las casillas correspondientes, derivadas de la tabla formada en la hipótesis de la movilidad perfecta.

La fórmula de la razón de movilidad, por tanto, es la siguiente:

$$R_{ij} = \frac{\text{valor observado}}{\text{valor esperado}} = n_{ij}\, N / n_{i.}\, n_{j}$$

El inconveniente principal de estas razones de movilidad consiste en que dependen de los marginales o totales.

En el caso de que la tabla de movilidad bruta coincida con la tabla de movilidad perfecta entonces los R_{ij} toman el valor de 1.

Sin embargo, cuando no sucede así, si $n_{i.} > n_{.j}$, entonces el máximo de $R_{ij} = N / n_{.j}$ y si $n_{.i} < n_{j.}$, entonces el máximo de $R_{ij} = N / n_{j.}$.

De la fórmula de independencia se deriva esta ecuación:

$$p_{11} = p_{1.}\, p_{.1}$$

o bien esta otra:

$$N\, n_{11} = n_{1.}\, n_{.1}$$

que es la misma que la anterior pero expresada en números absolutos.

Si esta ecuación se cumple en una tabla de movilidad, el status ocupacional del hijo no depende del status ocupacional del padre.

TABLA DE DIECISEIS CASILLAS

Es una tabla de contingencias de 4 filas por 4 columnas, que recibe el nombre de tabla de dieciséis casillas, por el número que tienen éstas. En esta tabla se relacionan dos variables en dos tiempos distintos T_1 y T_2. En ella se colocan horizontalmente las dos variables y sus dos dimensiones en el tiempo 2 y verticalmente dichas variables y dimensiones pero en el tiempo 1. En cada casilla se consigna la frecuencia conjunta que corresponde a las dimensiones de las variables que se cruzan en cada casilla.

Supongamos que resulta una tabla del tipo siguiente:

	amigos	*no amigos*	
En acuerdo político	100	100	
En desacuerdo	100	100	
	200	200	400

Supongamos que a este mismo grupo, pasados seis meses, se les vuelve a hacer otra encuesta y se relaciona nuevamente la amistad con el acuerdo político, resultando la tabla siguiente:

	amigos	*no amigos*
En acuerdo	150	50
En desacuerdo	50	150
	200	200

Cada una de estas tablas, consideradas independientemente, representa estáticamente la relación entre la amistad y el acuerdo político, en el momento dado, pero separadamente no nos dicen nada sobre la evolución en el tiempo de este grupo. En concreto nos dicen que en el tiempo 1 no había asociación entre la amistad y el acuerdo político, y en el 2, sí.

Ahora bien, si se comparan ambas tablas o se consideran no separadamente, sino conjuntamente, como una tabla única, esta tabla nos da ya una idea del proceso o transformación sufrido del tiempo 1 al tiempo 2 en el grupo.

En concreto, se puede deducir de ellas que las relaciones consonantes se han hecho más frecuentes en el tiempo 2 que lo eran en el 1.

Son consonantes las relaciones a la vez positivas o negativas en las dos variables, es decir, las de amigos en acuerdo y no amigos en desacuerdo, que han pasado en los dos casos de 100 en T_1 à 150 en T_2 .

Pero esta tabla sólo nos indica los cambios globales absolutos entre los dos tiempos, pero no todos los cambios que puede haber habido en el grupo.

Para conocer éstos es preciso formar la tabla de 16 casillas, que en este caso se supone es la siguiente:

		T_2				
		AI	*AI*	*AI*	*AI*	*Tot. T_1*
	AI	++	+-	- +	--	
T_1	++	50	20	10	20	100
	+-	30	20	0	50	100
	- +	50	0	40	10	100
	- -	20	10	0	70	100
Tot. T_2		150	50	50	150	400

Esta tabla se puede descomponer en los siguientes elementos: totales o marginales; casillas de la diagonal principal (4); casillas de la diagonal secundaria (4); y casillas laterales y centrales (8), en total 16.

Estos elementos presentan importantes propiedades que permiten el análisis cualitativo y cuantitativo de las variaciones sufridas de un tiempo a otro por las variables en estudio.

Las casillas de la diagonal principal (la que va de arriba a abajo) representan a los individuos que no han cambiado nada (ni en la amistad ni en las convicciones políticas) en los dos tiempos.

Por el contrario, las casillas de la diagonal secundaria (la que va de abajo a arriba), reflejan los individuos que han cambiado radicalmente (tanto en la amistad como en las convicciones políticas) de un tiempo a otro.

Las casillas laterales contienen los individuos que del tiempo 1 al tiempo 2 han cambiado respecto sólo a una de las variables, permaneciendo estables en cuanto a la obra.

ANALISIS CUANTITATIVO Y CUALITATIVO DE LAS TABLAS

Según lo anterior, sumando las cifras de las casillas de la diagonal principal, tendremos el número de los individuos que no han cambiado nada (en concreto en este ejemplo 180).

Sumando las de la diagonal secundaria, sacamos los que han cambiado radicalmente. En concreto, en este caso 40.

Sumando las restantes casillas laterales, los que han cambiado parcialmente se obtienen en total 180. De ellos 120 en la amistad y 60 en las ideas políticas.

El número total de cambios será pues igual a la suma de los cambios totales más la de los parciales, 40 y 180 = 220.

Vemos, por tanto, que mientras el número total de cambios absolutos en las tablas iniciales de los marginales era 100, el número efectivo y real de cambio se eleva a 220, según esta tabla.

Además, leyendo cada una de las casillas se puede concretar el sentido y la importancia de todos los cambios y no cambios ocurridos, así como se pueden comparar también todos entre sí.

También es conveniente distinguir las llamadas casillas fuertes de las casillas débiles. Las primeras son las que tienen una frecuencia relativa

elevada, en este caso por ejemplo 10 o más, y casillas débiles las de frecuencia baja o inferior a 10. (Ver ejercicios 243 a 247)

ANALISIS CUALITATIVO

El examen comparativo de los cambios habidos y de las casillas fuertes y débiles se puede concretar en proposiciones lógicas sobre el sentido y la importancia de la influencia de una variable sobre otra, que se llaman reglas secuenciales.

Cualquier casilla de la tabla que indique un cambio representa una secuencia o sea el paso de una posición en T_1 a otra en T_2.

Según esto se llama regla secuencial a cualquier proposición sobre la frecuencia de una sola secuencia, o sobre la importancia relativa de dos o más secuencias.

Por ejemplo, la amistad y las convicciones políticas tienden a coincidir. La no amistad y las divergencias políticas tienden a coincidir.

Los no amigos de la misma opinión política pasan a ser amigos. −, +, +, +, valor fuerte en esta casilla. La oposición en amistad y política tiende a mantenerse.

ANALISIS DE LA CAUSALIDAD

Cuando en el análisis de rotación se dice que una variable X es causa de otra Y, se quiere decir dos cosas:

1. Que la variable X origina o tiende a originar la variable Y, en mayor proporción que al revés.

2. Que la variable X tiende a producir la conservación de la relación inicial con la variable Y, en mayor proporción que al revés.

En consecuencia descubrir cuál de las dos variables es en el análisis de rotación causa de la otra equivale a saber cuál de ellas es más fuerte en el sentido de que origina y conserva en más alto grado la otra.

En nuestro caso se trataría de saber si es la amistad la que origina y conserva en más alto grado la actitud política o al revés.

Para saber esto existe la siguiente fórmula, que relaciona el número de ajuste y de pérdidas de coherencia de la variable amistad, que sobrepasan el número que se puede esperar, por efecto del azar.

$$I_{A,B} = \frac{8\left(\frac{\Delta_H}{N_H} + \frac{-\Delta_V}{N_V}\right)}{N}$$

Si el resultado es positivo la variable A es más fuerte que la B, y al contrario si es negativo.

Siendo:

Δ_H = El determinante de las casillas centrales de las filas.
N_H = Suma de las frecuencias de estas casillas.
Δ_V = Determinante de las cuatro casillas centrales de las columnas.
N_V = Suma de las frecuencias de estas casillas.
N = Suma total de los individuos de la tabla.
A = La primera variable, la intención de compra.
B = La segunda, la exposición a la publicidad.

(Ver ejercicio 248)

EJERCICIOS

239

Dada esta tabla:

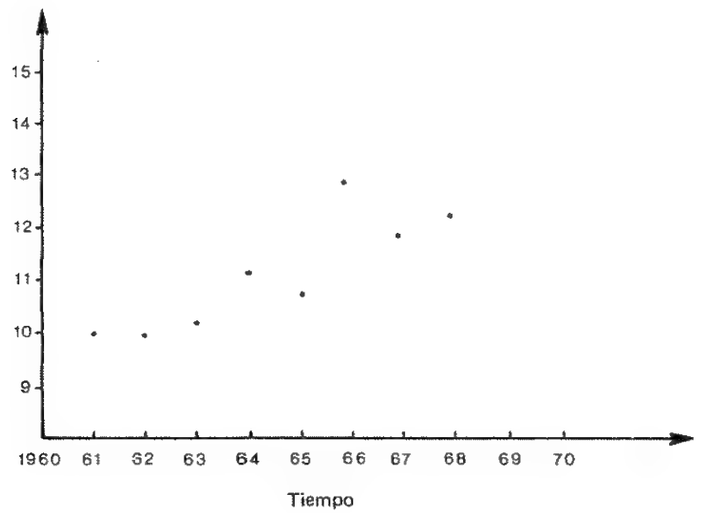
Porcentaje del total de presupuesto de gastos del Estado Español dedicado a Educación:

Años	%
1960	8.6
1961	9.7
1962	9.6
1963	8.9
1964	11.4
1965	10.6
1966	12.6
1967	11.5
1968	11.9
1969	14.7

Fuente: Informe Foessa de 1970, 1023.

A) Dibujar la huella empírica y especificar la ecuación que se crea puede ajustarse mejor a los datos.

RESPUESTA



Según la huella empírica se observa una tendencia general ascendente en la serie, pero con frecuentes altibajos dentro de esta tendencia general.

En ella no aparece sugerida una recta o un tipo determinado de curva de manera claramente identificable. De ahí que, dado que no se deriva del examen de la huella empírica una ecuación determinada, se puede intentar primero, como se suele hacer en estos casos, el ajuste de la ecuación más simple y sencilla, la lineal, que por otra parte parece puede ajustarse al conjunto de los datos mejor que ninguna otra.

B) *Estimar los parámetros de la ecuación lineal.*

RESPUESTA

Se puede realizar utilizando el procedimiento común de los mínimos cuadrados:

Los datos que se necesitan al efecto son:

Σy; ΣT; ΣT^2; ΣTy

Por tanto se puede formar la siguiente tabla de datos:

Σy	ΣT	ΣT^2	$\Sigma T \cdot Y$
8.6	0	0	0
9.7	1	1	9.7
9.6	2	4	19.2
8.9	3	9	26.7
11.4	4	16	45.6
10.6	5	25	53.
12.6	6	36	75.2
11.5	7	49	80.5
11.9	8	64	95.2
14.7	9	81	132.3
109.5	45	285	537.8

$$b = \frac{10\,(537.8) - (109.5)\,(45)}{10\,(285) - 45^2} = .546$$

$$a = \frac{109.5}{10} - . \ .546 \ \frac{45}{10} = 8.5$$

La ecuación resultante es:

$$Y = 8.5 + .546\ T$$

C) Evaluar el modelo.

RESPUESTA

Para evaluar el modelo se pueden hallar primero los valores estimados de la variable dependiente, porcentaje destinado a educación en el presupuesto de gastos, según la ecuación del modelo hallada y comparar estos valores estimados con los valores empíricos dados como datos en el modelo.

Después la evaluación se puede efectuar hallando el coeficiente de determinación entre las dos series de valores de la variable dependiente, empíricos y estimados según la ecuación, en cuanto dicho coeficiente representa la porción de la variación total de la variable dependiente Y, explicada por la ecuación hallada.

Valores de Y empíricos, Y	*Valores de Y estimados, $\hat{Y}$*	Y^2	$\hat{Y}^2$	$Y.\hat{Y}$
8.6	8.50	73.96	72.25	73.10
9.7	9.05	94.09	81.90	87.78
9.6	9.59	92.16	91.97	92.06
8.9	10.14	79.21	102.82	90.25
11.4	10.68	129.96	114.06	121.75
10.6	11.23	112.36	126.11	119.04
12.6	11.78	145.44	138.77	142.07
11.5	12.32	132.25	151.78	141.68
11.9	12.87	141.61	165.64	153.15
14.7	13.42	216.09	180.10	197.27
109.5	109.58	1217.13	1225.40	1218.15

$$r_2 = \frac{[10\,(1218.15) - (1095)\,109.58]^2}{[10\,(1217.13) - 1095.12^2]\,[10\,(1225.40) - 109.58^2]}$$

$$r^2 = .74$$

El coeficiente de determinación hallado, r = .74, indica el grado de conformidad entre las dos series de valores comparados, empíricos y estimados, y si bien no es muy elevada, se puede considerar aceptable.

240

Las siguientes tablas de movilidad procedentes de la obra de S. M. Lipset y R. Bendix, "Social mobility in industrial Society", Berkeley, The University of Carolina Press, se derivan de estudios realizados dentro de un período que va de 1940 a 1950.

Los autores distinguen tres categorías de ocupaciones: I, manual; II, no manual; III, agrícola. A efectos de este ejercicio se han reordenado, asignando el orden I, a la categoría agrícola, al II, a la manual, y el III, a la no manual.

Francia

Ocupación del padre	*Ocupación del hijo* I	II	III
I	916	167	206
II	63	343	219
III	100	200	809

Japón

Ocupación del padre	*Ocupación del hijo* I	II	III
I	731	322	409
II	60	441	247
III	58	243	853

Fuente: Thomas W. Pullum. Measuring occupational inheritance. N.Y. Elsevier, 1975, p. 98-9.

Establecer las matrices de inmovilidad, de cambio, residual y de movilidad neta, calcular los parámetros e interpretar los resultados comparativamente.

RESPUESTA

Matrices de inmovilidad:

Francia

	I	II	III
I	916		
II		343	
III			809

Japón

	I	II	III
I	731		
II		441	
III			853

Matrices de cambio: Se forman consignando en las casillas simétricas, en relación a la diagonal principal, la menor frecuencia de las simétricas en la matriz inicial.

Francia

	I	II	III
I	–	63	100
II	63	–	200
III	100	200	–

Japón

	I	II	III
I	–	60	58
II	60	–	243
III	58	243	–

Matrices residuales:

Francia

	I	II	III
I	–	104	106
II	–	–	19
III	–	–	–

Japón

	I	II	III
I	–	262	351
II	–	–	4
III	–	–	–

Movilidad neta:

Francia

Sumas o marginales de las ocupaciones

	Del padre	*Del hijo*	*Difer.*
I	1289	1079	–210
II	625	710	+85
III	1109	1234	+123
	3023	3023	0

Japón

Sumas o marginales de las ocupaciones

	Del padre	*Del hijo*	*Difer.*
I	1462	849	–615
II	748	1006	+258
III	1154	1509	+355
	3364	3364	0

Parámetros:

De inmovilidad:

Para Francia: 809 + 343 + 916 = 2068 / N = .68 (En relación a un N de 3023).

Para Japón: 853 + 441 + 731 = 2025 / N = .60 (N = 3364).

De movilidad:

Para Francia: 200 + 100 + 63 + 219 + 106 + 167 = 955/2023 = .32

Para Japón: 3364 – 1709 = 1655/3364 = .40

De movilidad ascendente:

Para Francia: 104 + 106 + 19 = 219/3023 = .075.

Para Japón: 262 + 351 + 4 = 617/3364 = .18.

Interpretación: Los parámetros y las tablas muestran que tanto en Francia como en Japón predomina la movilidad ascendente sobre la descendente como lo demuestran los índices hallados, los dos positivos, y las matrices residuales, en la que sólo aparecen cifras en las casillas de ascenso, es decir las situadas a la derecha de la diagonal principal.

De la comparación entre las dos naciones se deriva que es más fuerte proporcionalmente la inmovilidad en Francia que en el Japón y correlativamente la movilidad, en general, es mayor en el Japon que en Francia, así como que la movilidad ascendente es proporcionalmente más del doble superior en Japón que en Francia, índices .18 y .075.

241

Dadas las tablas utilizadas en el ejercicio anterior, hallar las tablas de movilidad perfecta, y las razones de movilidad Interpretar los resultados.

RESPUESTA

Tablas de movilidad perfecta:

	Para Francia				Para Japón		
	I	II	III		I	II	III
I	460	303	526	I	369	437	656
II	224	147	255	II	189	224	335
III	395	260	453	III	291	345	518
	1079	710	1234		849	1006	1509

Razones de movilidad:

	Para Francia				Para Japón		
	I	II	III		I	II	III
I	1.99	.55	.39	I	1.98	.77	.62
II	.28	2.33	.86	II	.32	1.99	.74
III	.25	.77	1.79	III	.20	.70	1.65

Interpretación

De las matrices de las razones de movilidad se deriva que, en cuanto en la diagonal principal de esta matriz aparecen valores superiores a la unidad en todos los casos, mientras que en las restantes casillas, que reflejan la movilidad, los valores que figuran en ellas son inferiores a la unidad. La movilidad social tanto en Francia como en Japón es inferior en todos los casos a la que se daría en un estado de movilidad perfecta.

Dado que los valores que aparecen en la matriz de Francia, en las casillas de la diagonal principal, son superiores a los que figuran en la matriz de Japón en las mismas casillas, se puede afirmar que Francia se aleja de la situación de movilidad perfecta en un grado más acusado que el Japón.

Por otra parte, el hecho de que las razones de movilidad que aparecen en las casillas a la derecha de la diagonal principal sean superiores o más elevadas que sus simétricas de la izquierda de dicha diagonal, confirma lo ya indicado de que en ambas naciones la movilidad ascendente supera a la descendente.

242

Dadas las tablas de movilidad en los ejercicios anteriores:

A) Hallar las matrices de proporciones y de transición.

RESPUESTA

La matriz de transiciones es una matriz $P = [p_{ij}]$ en la que los p_{ij} son igual a n_{ij}/N.

Por tanto tendremos las siguientes matrices de proporciones:

	Francia				*Japón*			
	I	II	III	Σ	I	II	III	Σ
I	.30	.06	.07	.43	.22	.10	.12	.44
II	.02	.11	.07	.20	.02	.13	.07	.22
III	.03	.07	.27	.37	.02	.07	.25	.34
Σ	.35	.24	.41	1.	.26	.30	.44	1.

La matriz de transición es una matriz $R = [r_{ij}]$ en la que r_{ij} se derivan de esta fórmula:

$$r_{ij} = \frac{n_{ij}}{n_{i.}} = \frac{p_{ij}}{p_{i.}}$$

Por tanto las matrices serán:

	Francia					*Japón*			
	I	II	III			I	II	III	
I	.697	.140	.163	1.	I	.50	.22	.28	1.
II	.100	.55	.35	1.	II	.08	.59	.33	1.
III	.081	.189	.73	1.	III	.05	.21	.44	1.

243

En una encuesta de panel sobre la intención de compra de TV en color realizada a una muestra de 200 personas, se obtuvieron los siguientes resultados en dos sondeos verificados en mayo y octubre de 1974.

	Compra de aparato de TV en color	
	En mayo	*En octubre*
Sí	80	120
No	120	80
TOTAL	200	200

Interpretar la tabla.

RESPUESTA

Esta tabla muestra simplemente la variación total experimentada en la muestra de mayo a octubre respecto a la compra de televisión en color. Dice que en octubre, el número de los que piensan comprar aparato de televisión en color, ha aumentado en 40 y correlativamente ha disminuido en 40 el número de los que no piensan comprarla, en relación a mayo. No se puede afirmar, según la tabla, que cuarenta individuos que no pensaban comprar televisión en mayo han cambiado de opinión en octubre, porque este número puede ser más elevado, en cuanto también puede haberse producido el fenómeno contrario de que algunos que pensaban comprarla en mayo hayan cambiado de propósito en octubre.

244

En la encuesta anterior se han analizado con más detenimiento los resultados recogidos y se ha comprobado que de los 80 que pensaban comprar la televisión en color en mayo, 30 seguían pensando comprarla en octubre y 50 habían cambiado de opinión y ya no pensaban comprarla. Asimismo, de los 120 que en mayo no querían comprar aparato de televisión, en octubre querían comprarlo 90 y 30 seguían sin querer comprarlo.

Se pide formar con estos datos una tabla de rotación en la que se reflejen las transformaciones indicadas experimentadas de mayo a octubre, indicando, primero, cuáles son las variables de la tabla y sus dimensiones y su forma de disposición en la tabla.

RESPUESTA

Se trata de una sola variable, intención de compra del aparato, con dos dimensiones, sí y no, pero considerada en dos tiempos distintos. Tiempos 1, mayo, y tiempo 2, octubre. Como los valores de la variable en cuestión son distintos en los dos tiempos se pueden formar con ellos una tabla de contingencia, en este caso llamada de rotación, en la que los valores del tiempo 1 deben aparecer horizontalmente y los del tiempo 2, verticalmente.

	Tiempo 2. Octubre		
TIEMPO 1. MAYO	*Intención de compra*	*No intención de compra*	TOTAL
Intención de compra	30	50	80
No intención de compra	90	30	120
TOTAL	120	80	200

245

Dada la tabla anterior se pide:

a) Determinar en qué sentido amplía esta tabla la información en relación a la anterior.

b) Interpretar el significado de los marginales o totales.

c) Señalar el significado de las diagonales principales y secundaria.

d) Indicar el total de los que no han cambiado de propósito de mayo a octubre y el de los que sí han cambiado y en qué sentido lo han hecho.

RESPUESTA

1. Esta tabla señala todas las variaciones de mayo a octubre en la intención de compra y el sentido de las mismas y no como la anterior sólo el aumento total de los que pensaban comprar el aparato de una fecha a otra.

2. Los marginales, 80 y 120; 120 y 80 reproducen la tabla del ejercicio anterior y tienen, por tanto, el mismo sentido que ella.

3. La diagonal principal señala los que no han cambiado de opinión de mayo a octubre. Treinta individuos que tenían intención de comprar en mayo la siguen teniendo en octubre e igualmente otros 30 que no tenían intención de comprar en mayo no siguen teniéndola en octubre.

Por el contrario, la diagonal secundaria indica las variaciones sufridas. Noventa individuos que no tenían intención de compra en mayo la tenían en octubre y, al revés, 50 que la tenían en mayo no la tenían en octubre.

4. El total de los individuos que no y sí han cambiado de intención de compra viene dado, en consecuencia, por la suma de los valores de las casillas de la diagonal principal y secundaria, respectivamente. Los que no han cambiado son, pues, 30 + 30 = = 60 y los que han cambiado 50 + 90 = 140.

246

En el estudio de panel anterior también se investigó en los dos tiempos en los individuos de la muestra los que habían estado sometidos a la publicidad para comprar el aparato de TV y los que no. La tabla derivada de los resultados obtenidos al tener en cuenta esta circunstancia fue la siguiente:

	TIEMPO 2				
Tiempo 1	*C P* ++	*C P* +−	*C P* −+	*C P* −−	*Total*
C P ++	25	10	5	10	50
C P +−	15	10	0	25	50
C P −+	25	0	20	5	50
C P −−	10	5	0	35	50
TOTAL	75	25	25	75	200

Signos:

C + = Intención de compra.
C − = No intención de compra.
P + = Exposición a la publicidad.
P − = No exposición de la publicidad

Se pide:

1. Formar las tablas que muestren la relación entre la publicidad y la intención de compra tanto en el tiempo 1 como en el tiempo 2, utilizando los marginales de la tabla.

2. Interpretar la relación entre ambas variables en los dos tiempos de la tabla.

RESPUESTA

TIEMPO 1

Intención de compra	*Publicidad* *Sí*	*No*
Sí	50	50
No	50	50

TIEMPO 2

Intención de compra	*Publicidad* *Sí*	*No*
Sí	75	25
No	25	75

En el tiempo 1 la tabla refleja que no hay asociación alguna entre la publicidad y la intención de compra, épsilon, 0. En cambio, en el tiempo 2 la tabla muestra una asociación apreciable, épsilon, 50 por 100 entre la publicidad y la intención de compra.

247

En la tabla de rotación del ejercicio anterior, determinar:

1. El número absoluto de cambios de mayo a octubre en relación a la compra y a la publicidad.

2. El número de los que en este período no han cambiado respecto a dichas variables.

3. El número de los que han cambiado radicalmente con referencia a la intención de compra y a la publicidad.

4. El número de los que han cambiado en parte, es decir, en relación a una sola de dichas variables.

5. El número total de cambios.

6. Las casillas fuertes de la tabla que no estén en la diagonal principal.

RESPUESTA

1. El número absoluto de cambios se deriva de la comparación de los marginales en el tiempo 1 y en el tiempo 2. Se tiene que los C + P + de 50 en el primer tiempo han pasado a 75 en el segundo, luego han aumentado en 25. En cambio, los C +, P − y los C − P + de 50 en el primer tiempo los dos han pasado en el segundo tiempo también los dos a 25; luego han perdido 25 cada uno. En cuanto a los C − P − de 50 en el primer tiempo han pasado a 75 en el segundo; por tanto, han aumentado en otros 25. Se ve que han descendido la cifra absoluta de las posiciones mixtas a favor de las posiciones homogéneas respecto a las dos variables. Este cambio se ha producido en la misma proporción.

2. El número total de los individuos que no han cambiado en nada de un tiempo a otro se halla, como se sabe, en la diagonal principal. Sumando sus valores se tiene 25 + 10 + 20 + 35 = 90. En relación al total de individuos de la muestra representa un porcentaje relativamente alto, el 45 por 100.

3. El número de los que han cambiado radicalmente respecto a las dos variables está en la diagonal secundaria. Suman 10 + 10 = 20. Los cambios radicales son iguales en los dos sentidos y representan un 10 por 100 del total.

4. El número de los que han cambiado parcialmente, en una sola de las variables, se encuentra en las restantes casillas, las no diagonales. Son en total 15 + 25 + 25 + + 5 + 10 + 5 + 5 + 0 = 90. De ellos, 60 han cambiado en la intención de compra, pero no en relación a la publicidad y 30 al revés.

5. El número total de cambios se puede obtener bien sumando los que han cambiado totalmente y en parte: 90 + 20 = 110, bien restando del total de individuos, 200, los que no han cambiado, 90.

6. Las casillas fuertes son la + − en el primer tiempo, con 15 individuos que han pasado en el segundo tiempo a + +, y también + − en el primer tiempo, con 25, que han pasado a − − en el segundo tiempo; así como la − +, con 25 en el primer tiempo, que ha pasado a + + en el segundo tiempo. Como se ve, todos los cambios importantes se han producido de posiciones mixtas a posiciones homogéneas o concordes en las dos variables. Predominan también los cambios en la intención de compra positivos, 40, sobre los negativos 25. En síntesis se puede decir que los cambios principales se han efectuado en la dirección siguiente: los que no tenían intención de comprar y estaban sometidos a la publicidad en el tiempo 1 han cambiado de actitud en el tiempo 2 y piensan comprar en este tiempo, mientras que los que en el tiempo 1 tenían intención de comprar y no estaban sometidos al efecto de la publicidad han pasado en el tiempo 2 a no tener intención de comprar.

248

Dados la tabla y el análisis descriptivo de los ejercicios anteriores 329 y 330, se pide analizar el sentido de la causalidad en el cambio producido en los dos tiempos, respecto a las dos variables, aplicando la fórmula siguiente:

$$I_{AB} = \frac{8\left(\frac{\Delta_H}{N_H} + \frac{-\Delta_V}{N_V}\right)}{N}$$

RESPUESTA

Aplicando la fórmula:

$$I_{AB} = \frac{8\,\frac{\Delta_H}{N_H} + \frac{-\Delta_V}{N_V}}{N_2 + 5}$$

se tiene:

$$I_{AB} = \frac{8 \times \frac{\begin{matrix}15 & 25\\ 25 & 5\end{matrix}}{70} + \frac{\begin{matrix}10 & 5\\ 5 & 0\end{matrix}}{20}}{200} = \frac{8 \times \left(-\frac{550}{70} + \frac{-25}{20}\right)}{200} = -0{,}36$$

El resultado es negativo, luego es más fuerte la variable B, exposición a la publicidad, que la variable A, intención de compra.

32. ANALISIS SOCIOMETRICO

NOCIONES BASICAS DE LA SOCIOMETRIA

El creador de la sociometría fue Jacob Levy Moreno, nacido en Bucarest, de origen judío, discípulo de Freud y colaborador de Jung. Iniciada en la década de 1920, fue desarrollada posteriormente por el fundador y sus discípulos y aplicada ampliamente en todo el mundo a partir de entonces en la investigación y en el trabajo social. En España, se pueden destacar los trabajos en este campo de Artur Arruga i Valeri, quien en su reciente tercera edición (1983) de su obra, *"Introducción al test sociométrico"*, ofrece una nueva versión muy completa y sistemática de esta técnica, por lo que será seguida especialmente en esta exposición.

El término sociometría se compone de *socio*, que procede del verbo latino *socio-as-are*, asociar, y éste de *socius*, con significación de compañero; y de *metría*, del griego *metron*, medida.

De acuerdo con este origen etimológico, en la sociometría se distinguen dos sentidos: uno amplio y otro estricto. En sentido amplio, el mismo Moreno define la sociometría (1975, 37) como "la ciencia que mide las relaciones interpersonales". En sentido estricto, la sociometría mide, no ya las relaciones interpersonales sin distinción y, por tanto, de cualquier clase que sea, sino dichas relaciones referidas concretamente a una situación de elección.

Entre las nociones básicas de la sociometría se pueden distinguir:

a) El grupo social;

b) El átomo social;

c) Las relaciones sociales; y

d) Los "teles" sociales.

Como grupo social se entiende el conjunto de personas que se *conocen* y se *relacionan* entre sí. Cada uno de los sujetos o elementos que forman el grupo social recibe el nombre de átomo social. Estos átomos están constituidos por cada uno de los individuos del grupo más el conjunto de relaciones sociales de las que forman parte.

Las relaciones sociales que unen a los miembros del grupo pequeño entre sí o con los miembros de otros grupos pueden ser: activas, pasivas o recíprocas; horizontales o verticales; y positivas o negativas.

Son activas aquéllas en las que el átomo social figura como sujeto emisor de la relación; pasivas, como sujeto receptor; y recíprocas, como emisor y receptor a la vez.

Las relaciones horizontales se dan entre átomos sociales del mismo nivel jerárquico y las verticales entre superiores e inferiores en una jerarquía social determinada.

Las relaciones sociales positivas implican elección, simpatía o atracción respecto a un átomo social, y las negativas, no elección, antipatía o rechazo.

Cada una de las relaciones positivas o negativas, de aceptación o rechazo, reciben el nombre genérico de "téle", término que designa específicamente la unidad de sentimiento de un sujeto respecto a otro.

EL TEST SOCIOMETRICO

La sociometría se funda en el instrumento de observación que Moreno llama el test sociométrico.

Consiste en un cuestionario de un reducido número de preguntas, al que Moreno asigna la finalidad de medir el grado de organización de un grupo.

Mayntz y otros (1975, 160) clasifican de la manera que se expone a continuación, las relaciones sociales que se pueden investigar mediante el test sociométrico.

1. Relaciones de atracción o rechazo. Se averiguan preguntando en el test sociométrico qué miembros del grupo se prefieren o, en su caso, a cuáles se estima menos. Se debe entender en general, es decir, sin particularizar motivos de atracción o rechazo.

2. Preferencias de interacción. Aquí se trata de averiguar los miembros del grupo con los que se desearía o no entrar en una relación determinada. Por ejemplo, con quién se preferiría dar un "golpe", realizar tal proyecto, formar equipo con un objetivo específico, etc. En este caso, en las contestaciones pueden influir otras consideraciones, distintas de la simpatía o antipatía, de las que es preciso ser consciente y tenerlas en cuenta en la aplicación del test.

3. Relaciones interactivas de hecho. En este caso, se preguntan relaciones efectivas. Por ejemplo, con quién se sale al cine, al baile, de "chateo", etc.

Por tanto, según el tipo de relaciones o criterios que se pretendan investigar, así se formularán las preguntas del test sociométrico. En ellas se tendrá cuidado especial de que no confundan a los encuestados y resulten efectivamente reveladoras de las relaciones que se indagan.

La preparación del test sociométrico debe basarse en un conocimiento fundamental previo del grupo, respecto a aspectos tales como:

- Su tamaño, que influye en el carácter más o menos compacto o diluído de las relaciones del grupo.
- Antigüedad del grupo, factor del que depende la extensión y profundidad de las relaciones entre los sujetos del grupo.
- Fecha de incorporación de sus miembros, como indicador de su mayor o menor integración en el grupo.
- Objetivos del grupo, que condicionan los objetivos del test.

Supuesto este conocimiento previo del grupo, en la preparación del test sociométrico se pueden distinguir las siguientes operaciones básicas:

1. Determinación de los objetivos o fines del test sociométrico y del tipo concreto de criterios o relaciones sociométricas que se piensa investigar mediante aquél.

2. Definir la situación preferencial decidiendo:

a) Si se va a permitir que las elecciones y rechazos se extiendan únicamente a los miembros del grupo investigado o también a sujetos extraños a él.

b) Si el test se va a limitar a las elecciones o incluirá además los rechazos. Aunque la inclusión de estos últimos puede suponer inconvenientes psicológicos y técnicos adicionales, su introducción es

siempre necesaria, si se quiere obtener un conocimiento preciso del grado de integración y de las tensiones del grupo, así como de las causas de su mal funcionamiento.

c) Si se va a investigar o no la percepción sociométrica, es decir, la opinión de cada sujeto del grupo sobre quiénes le parece que le han escogido o rechazado. Su inclusión puede revelar la empatía, o capacidad de adivinación de los sentimientos de los demás hacia uno mismo.

d) El número de elecciones y, en su caso, rechazos. Este puede ser libre o limitado. La no limitación hace más complejo sobre todo el sociograma, pero da una imagen más completa de la expansividad del grupo, del rango sociométrico de los individuos y de las interrelaciones entre sus miembros.

e) Orden de las elecciones y rechazos. Cuando un elemento del grupo elige o rechaza a más de otro elemento del grupo, es recomendable que se consignen por orden, empezando por el más preferido o rechazado y terminado por el menos. La cuestión que se plantea entonces es cómo valorar este orden. Un procedimiento admitido es asignar cinco puntos a la primera elección o rechazo; cuatro a la segunda; tres, a la tercera; dos, a la cuarta; y 1, a la quinta y restantes.

3. Redacción de las preguntas del test. Se realizará en consonancia con los objetivos fijados y las relaciones sociométricas que se ha decidido investigar.

Estas preguntas serán directas, sencillas y claras y, en el caso de que el test abarque también los rechazos y la percepción de las elecciones y rechazos, se referirán, al menos, a los cuatro puntos siguientes:

- Simpatías, preferencias o elecciones;
- Rechazos.
- Quién o quiénes se piensa nos han elegido, y
- Quién o quiénes se cree nos han rechazado.

La administración del test sociométrico se realiza estando el grupo reunido y cumplimentándolo individualmente cada sujeto con la supervisión del investigador, quien se estima debe no ser un desconocido para el grupo y gozar de su confianza.

Para que los miembros del grupo cooperen adecuadamente, deben ser motivados mediante la explicación de los objetivos del test y del interés que reviste para ellos.

Es obvio que la contestación al test debe ser estrictamente personal. Para lograr este requisito, no se debe permitir comunicación alguna entre los miembros del grupo investigado durante la administración del test y se garantizará el secreto de las respuestas y de las elecciones.

EL ANALISIS SOCIOMETRICO

El test sociométrico es, como se ha indicado, el instrumento de observación utilizado en la sociometría. Una vez aplicado al grupo investigado, los datos obtenidos mediante él deben ser objeto de análisis. Así como en la encuesta, la base de este análisis es la agrupación y clasificación en tablas de los datos obtenidos, similarmente en la sociometría la base son las sociomatrices, que agrupan y clasifican los datos proporcionados por el test sociométrico.

En los primeros tiempos de la sociometría, este análisis se centró en el sociograma, que es la representación gráfica de las elecciones y rechazos, en su caso, de un test sociométrico. En la actualidad, al contrario, se basa en la sociomatriz, habiendo quedado relegado el sociograma a un segundo plano complementario.

Reciben el nombre de sociomatrices, las matrices o tablas que recopilan los resultados del test sociométrico. Tales matrices son conjuntos de filas y columnas de números, que, en este caso particular, son cuadradas, es decir, con el mismo número de filas y columnas, cada una de las cuales corresponde a un individuo de los que forman el grupo al que se ha aplicado el test sociométrico. Las filas y columnas se pueden titular con el número asignado a cada individuo en el grupo o con las siglas de sus nombres respectivos.

En las casillas de la matriz se consignan con un 1, las elecciones, y con un 1 de distinto color, los rechazos, realizados por el individuo al que se refiere el número o sigla de la fila correspondiente y con un 0 las no elecciones ni rechazos. Cuando se valoran éstos, entonces en lugar del 1, se colocará en cada casilla de la matriz la cifra que corresponda según el orden de las elecciones hechas.

Como las elecciones y rechazos se han de situar en la columna del sujeto elegido o rechazado, las columnas de la matriz reflejan, por tanto, las elecciones y rechazos de que haya sido objeto el individuo al que se refiera la columna en cuestión.

La percepción de las elecciones y rechazos se pueden consignar en la sociomatriz mediante signos especiales, e igualmente las elecciones recíprocas.

VALORES SOCIOMETRICOS

La sociomatriz, según se acaba de indicar, es una tabla de datos. Supuesto esto, los valores sociométricos no son otra cosa que las sumas o totales de los distintos tipos de datos que se recogen en la sociomatriz.

Con Arruga i Valeri se pueden distinguir los valores sociométricos principales siguientes:

- Sp: Rango, o en inglés *status*, positivo = Número de elecciones positivas que se reciben.
- Sn: Rango o *status* negativo = Número de rechazos obtenidos.
- Si se trata de elecciones y rechazos valorados, entonces se representan por Sp val y Sn val.
- Ep y En: Expansividad positiva y negativa = al número de elecciones positivas o negativas que hace cada sujeto.
- Rp y Rn: Elecciones recíprocas positivas o negativas.
- OS: Oposición de sentimientos = número de elecciones de cada sujeto que se corresponden con rechazos y al revés.
- Pp y Pn: Percepción positiva o negativa de elecciones y rechazos = al número de los que cada sujeto cree que le eligen o rechazan.
- PAp y PAn: Número de percepciones de elecciones y rechazos acertadas.
- FP: Falsa percepción o contraste entre elecciones y rechazos positivos y la percepción de éstos.

INDICES SOCIOMETRICOS

Son coeficientes numéricos que cifran cuantitativamente diversas propiedades y características del grupo. Se pueden dividir en individuales, referentes a los individuos en particular, y grupales, al grupo en su conjunto.

Entre los primeros están los siguientes:

Conexión afectiva: CA. Mide la proporción de las relaciones recíprocas de un sujeto, respecto al número de elecciones que dicho sujeto recibe.

Fórmula:

$$CA = Rp/Sp$$

El máximo es 1 cuando todas las elecciones que recibe un sujeto son recíprocas.

Atención perceptiva positiva y negativa, APp y APn. Es la relación entre el número acertado por el sujeto de elecciones positivas o negativas y el número total de elecciones o rechazos que recibe, respectivamente.

Fórmulas:

$$APp = PAp/Sp;\ APn = PAn/Sn$$

El máximo es también 1 y tiene lugar cuando coinciden las percepciones y las elecciones efectivas.

Realismo perceptivo positivo y negativo, RPp y RPn. Relaciona las percepciones acertadas positivas o negativas con el total de percepciones positivas o negativas, respectivamente, del sujeto, según las siguientes fórmulas:

$$RPp = APp/Pp \text{ y } RPn = Apn/Pn$$

Su máximo es 1 cuando ambas percepciones coinciden.

Indices de distancia y rango sociométricos. El primero se forma reuniendo los datos de elecciones y rechazos y percepciones de ambos, de cada individuo del test sociométrico respecto a todos los demás.

La distancia entre dos individuos A y B se halla del siguiente modo:

Con un punto:

- La elección de A por B;
- La elección de B por A;
- La percepción de elección de B por A, y
- La percepción de elección de A por B.

Con un cero:

- La no elección ni rechazo y la no percepción de elección ni de rechazo de A por B y de B por A.

Con menos un punto:

– El rechazo tanto de B por A como el de A por B.

– La percepción de rechazo de A por B y la de B por A.

La distancia entre dos sujetos, cuando se tienen en cuenta las elecciones y rechazos y las percepciones de ambos, puede oscilar entre 4, que designa la máxima unión y –4, la máxima distancia.

Si están valoradas las elecciones y rechazos en la forma indicada, entonces, en lugar de 1 punto, se asignan a las elecciones y rechazos los puntos que correspondan a la valoración adoptada. En este caso, el índice de distancia varía de 12 a -12.

El índice de *status* sociométrico se halla dividiendo la suma de las distancias de cada individuo a todos los demás por el número de sujetos del grupo N, menos 1. Los valores máximos y mínimos de este índice son también 4, –4, 12 y –12.

Como índices grupales se pueden mencionar:

Indice de expansividad del grupo. Es igual al cociente entre el número total de elecciones hechas por el grupo y N. Puede ser positivo, si sólo tiene en cuenta las elecciones; negativo, si sólo los rechazos; y total, si ambos.

Fórmulas:

$$IEp = \Sigma Sp/N;\ IEn = \Sigma Sn/N;\ IEt = (\Sigma Sp + \Sigma Sn)/N$$

Máximo: N–1

Indice de coherencia. Compara el total de elecciones recíprocas del grupo, Σ Rp, con el total de elecciones, recíprocas o no.

Fórmula:

$$Icoh = \Sigma Rp/\Sigma Sp.$$

Máximo: 1

Indice de cohesión. Su fórmula es la siguiente:

$$Ic = \frac{\text{Número de elecciones mutuas del grupo}}{\text{Número total de posibles elecciones mutuas}}$$

Este último es igual al número máximo posible de elecciones del grupo, o sea N. (N-1), partido por dos. El mayor valor de este índice es 1 y tiene lugar cuando todos los miembros del grupo se eligen unos a otros, y, por tanto, se eligen recíprocamente.

Indice de disociación. Es opuesto al anterior. Consiste en el número de rechazos recíprocos, partido por el número máximo de rechazos posibles.

Fórmula:

$$ID = \frac{Rn}{(N)\,(N-1)/2}$$

Máximo: 1.

Indices de atención y realismo perceptivos positivos y negativos del grupo. Tienen la misma significación respecto al grupo que los individuales.

Fórmulas:

$$APpg = \frac{\Sigma\ Pp}{\Sigma\ Sp};\qquad APng = \frac{\Sigma Pn}{\Sigma Sn}$$

$$RPpg = \Sigma APp/\Sigma Pp;\ RPng = \Sigma APn/\Sigma Pn$$

Indice de preferencias del subgrupo. Si se distinguen en un grupo dos subgrupos, por ejemplo, chicos y chicas, este índice compara las elecciones dirigidas a miembros del subgrupo propio con las dirigidas a los miembros del otro subgrupo.

Fórmula:

$$IPS = \frac{I_1\ (N - N_1)}{E_1\ (N_1 - 1)}$$

Donde I_1 y E_1 son la suma de elecciones hechas dentro y fuera del subgrupo, respectivamente. Si es mayor que 1, indica que los miembros del subgrupo se prefieren entre sí más que a los miembros del otro subgrupo.

Indice de unión del subgrupo. Refleja la integración o unión del subgrupo entre sí y respecto al otro subgrupo.

Fórmula:

$$US_1 = \frac{I_1 \ \Sigma Sp}{Ep_1 \ \Sigma Sp_1}$$

donde Ep_1 y Sp_1 son el número de elecciones positivas hechas y recibidas por el subgrupo, respectivamente. Si es mayor que 1 indica que el subgrupo está muy unido entre sí y aislado del otro subgrupo; si es menor, lo contrario.

EL SOCIOGRAMA

El test sociométrico sirve también para construir el sociograma, que representa gráficamente la estructura de las relaciones existentes en un grupo en un momento dado. Permite abarcar con una ojeada las líneas de comunicación del grupo y los esquemas de atracción y repulsión.

En realidad, un sociograma es un grafo orientado o dirigido, de ahí que se pueda aplicar y de hecho se aplique a su estudio, sobre todo en el aspecto de las redes de comunicación, la teoría matemática de los grafos que forma parte de la topología, nueva rama de las matemáticas modernas.

Para construir el sociograma se numeran correlativamente al azar todas las respuestas al test de los miembros del grupo, en las que deben consignar su nombre.

Hecha esta operación, se confecciona una lista en la que se hace figurar primero el número del que ha dado la respuesta y después, en columnas separadas, los números de los que ha elegido y rechazado.

En el sociograma se indica cada sujeto por su número rodeado de un círculo. Si hay chicos y chicas se pueden distinguir, por ejemplo, asignando a unos un círculos y a otras un triángulo. Las elecciones y rechazos se consignan mediante flechas de trazo continuo o discontinuo, respectivamente, o de distinto color, en las que la punta de la flecha termina en el elegido o rechazado. Si la elección es mutua, o bien se dibujan dos flechas, o bien una sola, con puntas en los dos extremos.

A veces puede ser conveniente construir, no un sólo sociograma, sino los siguientes separados:

- De elecciones recíprocas;
- De rechazos recíprocos;
- De elecciones en general; y
- De rechazos en general.

Para facilitar la construcción del sociograma existe la técnica de Northway (1952), que consiste en dibujar tres círculos concéntricos, colocando en el primero los individuos con un número de elecciones significativamente alto; en el tercero, los significativamente bajos; y en el medio, los no significativos. En el sociograma y sociomatriz se pueden distinguir las siguientes situaciones:

- Elecciones unilaterales.
- Elecciones recíprocas, o parejas.
- Cadenas. Serie de sujetos en el que el primero, elige al segundo, éste al tercero y así sucesivamente.
- Cliques. Clique es una palabra inglesa que significa peña, pandilla. En sociometría designa un conjunto de individuos, que se eligen todos recíprocamente entre sí.
- Aislados. Ni eligen, ni son elegidos.
- Marginados, eligen, pero no son elegidos.
- Independientes, los eligen, pero no eligen ellos.
- Populares, los muy elegidos por los demás.
- Rechazados, los que reciben muchos rechazos.

ANALISIS INDIVIDUAL Y DEL GRUPO EN SU CONJUNTO

El primero se debe realizar mediante el examen comparativo de los valores e índices sociométricos y la construcción del sociograma individual.

Este sociograma sitúa en el centro al individuo de que se trate y a los demás individuos en círculos concéntricos sucesivos, más o menos alejados según la distancia sociométrica a que se encuentren del sujeto analizado.

El análisis del grupo en su conjunto, se debe efectuar mediante la consideración comparativa de los totales de los valores sociométricos individuales y atendiendo, de modo especial, a los índices sociométricos globales que reflejan la expansividad del grupo, su coherencia, asociación, disociación, atención y realismos perceptivos y la preferencia y unión interna de los subgrupos distinguidos en el subgrupo.

También se puede obtener información sobre el grupo a través del examen del sociograma y la sociomatriz y la aplicación del cálculo ma-

tricial, sobre todo respecto a las distintas situaciones sociométricas antes señaladas.

Es preciso señalar que los resultados de ambos análisis se han de interpretar, teniendo en cuenta los datos de experiencia o de otro tipo que ya se posean de los individuos y del grupo.

EJERCICIOS

249

Se ha aplicado un test sociométrico, sin limitación de elecciones y de rechazos y por orden de ambos, a un grupo de seis estudiantes, tres chicos, números 1, 2 y 3, y tres chicas, números 4, 5 y 6, en el que el criterio de las elecciones y rechazos fue la formación de nuevas clases en el curso próximo, obteniéndose los siguientes resultados:

N°	*Elige*	*Rechaza*	*Cree que lo eligen*	*Cree que lo rechazan*
1	2, 6	3, 4	2, 6	3
2	1, 5, 4	6	1,4	3
3	4, 5, 6	1, 2	, 5	1,2,6
4	1, 2, 3, 5	6	3,5	6
5	4	6	3,4,2	— —
6	1, 2, 3	4, 5	1,	, 4

Formar la sociomatriz correspondiente:

RESPUESTA

	1	2	3	4	5	6
1	0	$\hat{5}$	x ⑤	④	0	$\hat{4}$
2	$\hat{5}$	0	x $\hat{0}$	$\hat{3}$	4	⑤
3	x ⑤	x ④	0	5	$\hat{4}$	x 3
4	5	4	$\hat{3}$	0	$\hat{2}$	x ⑤
5	0	$\hat{0}$	$\hat{0}$	$\hat{5}$	0	⑤
6	$\hat{5}$	$\hat{4}$	3	x ⑤	④	0

Signos:
0 = No elección ni rechazo
Números sin círculo = Elecciones valoradas
Números con círculo = Rechazos valorados
ˆ = Percepción de elecciones
x = Percepción de rechazos.

250

Dada la sociomatriz anterior, hallar sus valores sociométricos.

RESPUESTA

Su signo y significación se indica en la página 684. Se hallan sumando simplemente el número de los distintos tipos de datos de la sociomatriz.

A este respecto se indica, por los signos siguientes:

— , las elecciones recíprocas
~ los rechazos recíprocos, y
φ la oposición de sentimientos

	1	2	3	4	5	6		Ep	En	Rp	Rn
1	0	$\hat{5}$	x ⑤ ~	φ 4	0	$\hat{4}$ —		2	2	2	1
2	$\hat{5}$	0	x $\hat{0}$	$\hat{3}$	4	⑤φ		3	1	2	0
3	x ⑤ ~	x ④	0	5	$\hat{4}$	x 3		3	2	2	1
4	5φ	4	$\hat{3}$	0	$\hat{2}$	x ⑤ ~		4	1	3	1
5	0	$\hat{0}$	$\hat{0}$	$\hat{5}$	0	⑤ ~		1	1	1	1
6	$\hat{5}$	4φ	3	x ⑤ ~	④ ~	0		3	2	2	2
Spral	15	13	6	13	10	7	64	16	9	12	6
Sp	3	3	2	3	3	2	16				
Snval	5	4	5	9	4	15	42				
Sn	1	1	1	2	1	3	9				
Pp	2	2	2	2	2	2	12				
PAp	2	1	1	2	2	1	10				

Pn	1	1	2	1	0	2	7
PAn	1	1	1	1	0	1	5
Fp	0	0	1	0	0	2	3
Os	1	1	0	1	0	1	4

251

Dada la sociomatriz formada en el ejercicio, hallar los índices sociométricos de cada individuo siguiente.

a) De conexión afectiva, CA.

RESPUESTA

Aplicando la fórmula CA = Rp/Sp

Núm. 1 : 2/3 = .66
Núm. 2 : 2/3 = .66
Núm. 3 : 2/2 = 1
Núm. 4 : 3/3 = 1
Núm. 5 : 1/3 = .33
Núm. 6 : 2/2 = 1

b) De atención perceptiva positiva, APp, y negativa, APn.

RESPUESTA

Aplicando la fórmula APp = PAp/Sp; APn = PAn/Sn

PAp	PAn
Núm. 1 : 2/3 = .66	Núm. 1 : 1/1 = 1
Núm. 2 : 1/3 = .33	Núm. 2 : 1/1 = 1
Núm. 3 : 1/2 = .50	Núm. 3 : 1/1 = 1
Núm. 4 : 2/3 = .66	Núm. 4 : 1/2 = .50
Núm. 5 : 2/3 = .66	Núm. 5 : 0/1 = 0
Núm. 6 : 1/2 = .50	Núm. 6 : 1/3 = .33

c) De realismo perceptivo positivo, RPp, y negativo, RPn.

RESPUESTA

Aplicando las fórmulas, RPp = PAp/Pp y RPn = PAn/Pn, se tiene:

RPp	RPn
Núm. 1 : 2/2 = 1	Núm. 1 : 1/1 = 1
Núm. 2 : 1/2 = .50	Núm. 2 : 1/1 = 1
Núm. 3 : 1/2 = .50	Núm. 3 : 1/2 = .5
Núm. 4 : 2/2 = 1	Núm. 4 : 1/1 = 1.0
Núm. 5 : 2/2 = 1	Núm. 5 : 0/0 = 1
Núm. 6 : 1/1 = .50	Núm. 6 : 1/2 = .50

d) Hallar la distancia sociométrica de cada individuo a todos los demás.

RESPUESTA

D/S del núm. 1

Al	Sp	Sn	Pp	Pn	T
2	10	0	2	0	12
3	0	-10	0	-2	-12
4	5	-4	0	-0	1
5	0	0	0	0	0
6	9	0	2	0	11
					12

Del núm. 4

Al	Sp	Sn	Pp	Pn	T
1	7	0	0	-1	6
2	5	-4	0	0	1
3	8	0	2	0	10
5	7	0	2	0	9
6	0	-10	0	-2	-12
					14

Del núm. 2

Al	Sp	Sn	Pp	Pn	T
1	10	0	2	0	12
3	0	-4	0	-2	-6
4	7	0	0	-1	6
5	4	0	1	0	5
6	4	-5	1	0	0
					17

Del núm. 5

Al	Sp	Sn	Pp	Pn	T
1	0	0	0	0	0
2	4	0	1	0	5
3	4	0	2	0	6
6	7	0	2	0	9
6	0	-9	0	-0	-9
					11

Del núm. 3

Al	Sp	Sn	Sp	Pn	T
1	0	-4	0	-1	-5
2	0	-2	0	-1	-3
4	8	0	2	0	10
5	4	0	2	0	6
6	6	0	0	-1	5
					13

Del núm. 6

Al	Sp	Sn	Pp	Pn	T
1	9	0	2	0	11
2	4	-5	1	0	0
3	6	-0	0	-1	5
4	0	-10	0	-2	-12
5	0	-9	0	0	-9
					-5

e) Hallar el índice de status sociométrico.

RESPUESTA

Resulta de dividir la distancia total por N – 1. Se tiene pues.

Núm. 1 : 12/5 = 2,40 ; Núm. 2 : 17/5 = 3,4
Núm. 3 : 15/5 = 3 ; Núm. 4 : 14/5 = 2,80
Núm. 5 : 11/2 = 2,20 ; Núm. 6 : -5/5 = -1

252

Dada la sociomatriz formada en el ejercicio, hallar los índices sociométricos grupales.

A) Indice de cohesión, IC:

RESPUESTA

Aplicando la fórmula:

$$IC = \frac{\Sigma Rp}{N(N-1)} = \frac{12}{6(6-1)} = \frac{12}{30} = .40$$

B) Indice de disociación, ID.

RESPUESTA

Fórmula:

$$ID = \frac{\Sigma Rn}{N(N-1)} = \frac{6}{30} = .20$$

C) Indice de coherencia.

RESPUESTA

Fórmula:

$$ICoh = \frac{\Sigma Rp}{\Sigma Sp} = \frac{12}{16} = .75$$

D) Indice de expansividad, positiva y negativa y mixta.

RESPUESTA

Fórmula:

$$IEp = \frac{\Sigma Sp}{N} = \frac{16}{6} = 2.66$$

$$IEn = \frac{\Sigma Sn}{N} = \frac{9}{6} = 1.50$$

$$IEt = \frac{\Sigma Sp + \Sigma Sn}{N} = \frac{25}{6} = 4,16$$

E) Indices de preferencias del subgrupo 1 y 2.

RESPUESTA

Fórmula:

$$Ipsi = \frac{Ii\,(N - Ni)}{Ei\,(Ni - 1)}$$

$$Ips_1 = \frac{2(6-3)}{6(3-1)} = .50; \quad Ips_2 = \frac{2(6-3)}{6(3-1)} = .50$$

H) *Indice de unión de los subgrupos 1 y 2.*

RESPUESTA

Fórmula:

$$VSi = \frac{Ii\ \Sigma Sp}{Epi\ \Sigma Spi}$$

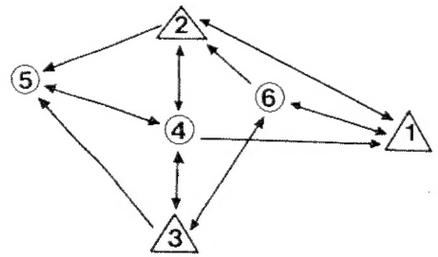
$$VS_1 = \frac{2(16)}{8(8)} = .50; \quad VS_2 = \frac{2(16)}{8(8)} = .50$$

253

Dados los resultados del test sociométrico iniciales, formar:

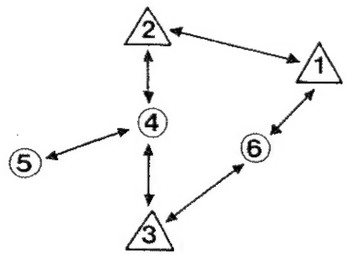
a) El sociograma de elecciones.

RESPUESTA



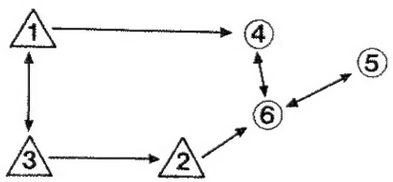
b) El sociograma de elecciones recíprocas:

RESPUESTA



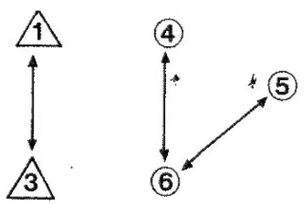
e) El sociograma de rechazos:

RESPUESTA



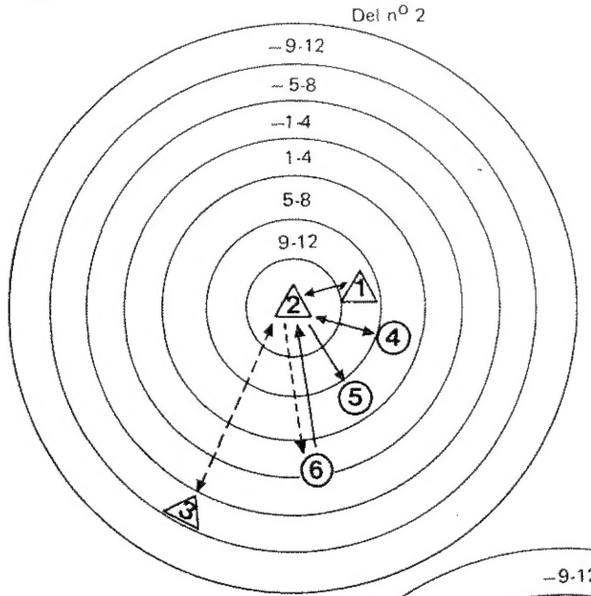
d) El sociograma de rechazos recíprocos:

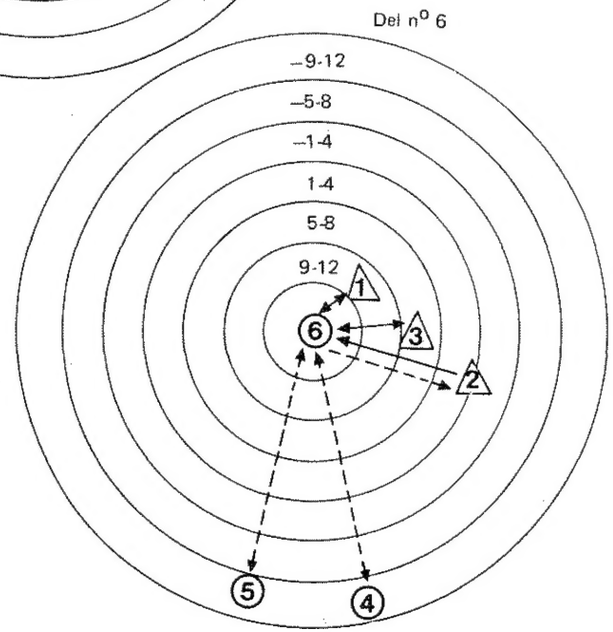
RESPUESTA



e) Los sociogramas individuales de los números 2 y 6.

RESPUESTA





254

Realizar los siguientes análisis:

a) Individual de los números 2 y 6.

RESPUESTA

Se puede formar primero un cuadro comparativo de sus valores e índices individuales:

N°	Sp	Spv	Sn	Snv	Pp	PAp	Pn	PAn	F	Os	CA	APp	APn	Rpp	RPn	Ds	SS
2	3	13	1	4	2	1	1	1	0	1	.66	.66	1	.5	1	17	3
6	2	7	3	15	2	1	2	1	2	1	1	.50	.33	.5	.5	-5	-1

La comparación de ambas series muestra que los valores e índices positivos del 2 son superiores a los del 6, salvo el de conexión afectiva, e inversamente en los negativos.

Concretamente, el número 2 es uno de los muchachos más populares del grupo. Su Sp, junto con el del 1 y el 4, es máximo. Su atención y su realismo perceptivos negativos son totales, pero no así los positivos, pues cree que le va a elegir el 5 y se equivoca. En sus relaciones positivas claramente predominan las intersexuales, con chicas, si bien le une una relación recíproca fuerte con otro chico, el 1, correspondida por éste. En cambio, otro chico, el 3, le rechaza. Con la chica 6 mantiene una oposición de sentimiento, pues ella lo elige a él y él la rechaza. Aunque es el individuo con status sociométrico más elevado en el grupo, sin embargo, no es totalmente aceptado por el grupo, ni él acepta a todos los miembros del grupo.

En cuanto a la número 6, es la más rechazada del grupo y la única que tiene una distancia sociométrica negativa respecto a todo el grupo y, por tanto, un *status* sociométrico negativo en él. Sus relaciones positivas son exclusiva y totalmente heterosexuales; elige a todos los chicos y su elección es correspondida, aunque no en primer lugar, por el 1 y por el 3, si bien es rechazada por el 2. En cambio está unida por relaciones negativas recíprocas con las otras dos chicas del grupo. No acierta en su percepción de los que la aceptan y rechazan en el grupo; incluso en la falsa percepción alcanza un valor de dos, lo que indica que en dos casos ha ocurrido de hecho lo contrario de lo percibido por ella. Por tanto, es la menos realista del grupo respecto a su posición en él. Se trata, pues, de una chica causa de división en el grupo, aceptada relativamente por el subgrupo de chicos y rechazada totalmente por el de chicas.

b) Del grupo en su conjunto.

RESPUESTA

La expansividad total del grupo es muy alta, 4,16, en relación al máximo 5. Dentro de ella es superior la expansividad positiva a la negativa. También es alto el realismo positivo y negativo del grupo: el primero 10/12 y 5/7 el segundo. Se puede inferir la hipótesis de que los miembros del grupo se conocen bien entre sí, por ser acaso un grupo antiguo.

El índice de cohesión del grupo, es decir, de conexiones recíprocas que son las más consistentes, no es alto, 0,40, en relación a un máximo posible de 1, pero es el doble que el índice contrario, el de disociación, 0,20. El que sí se puede considerar en principio alto, es el índice de coherencia positivo y negativo o proporción de las elecciones y rechazos recíprocos en relación al total efectivo de elecciones y rechazos: 0,75 el primero y 0,66 el segundo, respecto también de un máximo de 1.

No obstante, para un juicio más acertado de estos índices sería necesario poseer como patrón de comparación, información sobre los índices que normalmente presentan los grupos obtenidos de la misma población.

Lo que sí son bajos, son los índices de preferencia y de unión de los subgrupos, 0,50 los dos, e indican claramente que en cada subgrupo predominan las preferencias hacia y a la unión con los miembros del otro subgrupo y no con los del propio.

Esto se ve gráficamente en los sociogramas de rechazos y elecciones recíprocas. Según el primero, el grupo está partido por dos enfrentamientos: unos entre los chicos 1 y 3, y otro entre las chicas 4 y 5 con la 6.

Según el segundo, salvo la unión recíproca entre los chicos 1 y 2 y las chicas 2 y 5, todas las demás 5, son entre personas de distinto sexo.

En síntesis, se trata de un grupo muy activo sociométricamente, pero no muy integrado en su conjunto, sino con fuertes enfrentamientos internos dentro de los subgrupos que lo componen. Se podría formular la hipótesis de que son amorosos las relaciones intersexuales y que los enfrentamientos internos de los subgrupos se deben a rivalidades de este tipo.